Chromatin Compaction in Cornelia de Lange Syndrome

Emily Pritchard

For the Degree of Doctor of Philosophy

University of Edinburgh

2011

Contents

List of figures and tables

1 Introduction

2 Materials and methods

3 General effects of CdLS on the nucleus

4 Chromatin condensation at specific regions in CdLS

List of Abbreviations

2D: Two dimensional

3C: Chromatin conformation capture

3D: Three dimensional

4C: Circular 3C, or 3C on chip

5C: 3C carbon copy

6C: Combined 3C ChIP cloning

A: Amperes

ACH: Active chromatin hub

ANOVA: Analysis of variance

APC/C: Anaphase promoting complex

APO: Apolipoprotein

ATP: Adenosine tri-phosphate

Bio-16-dUTP: Biotin-16-dUTP

bp: Base-pairs

BrdU: Bromodeoxyuridine

BSA: Bovine serum albumin

CE: Centrifugal elutration

CdLS: Cornelia de Lange Syndrome

ChIP: Chromatin Immunoprecipitation

CNS: Central nervous system

CT: Chromosome territory

CTCF: CCCTC-binding factor

Dam: DNA adenine methyltransferase

dBBr: Dibromobimane

DCC: Dosage compensation complex

DEPC: Diethylpyrocarbonate

DIG-11-dUTP: Digoxigenin-11-dUTP

dH_2O: Distilled water

DMEM: Dulbecco's Modified Eagle Medium

DMR: Differentially methylated region

DMSO: Dimethyl sulphoxide

DNA: Deoxyribonucleic acid

DNMT: DNA methyltransferase

dNTP: Deoxyribonucleotides

DSB: Double strand break

DTT: Dithiothreitol

EBV: Epstein-Barr virus

ECL: Enhanced chemiluminescent

EDC: Epidermal differentiation complex

EDTA: Ethylenediaminetetraacetic acid

ENU: *N*-ethyl-*N*-nitrosourea

ER: Oestrogen receptor

ES: Embryonic stem

EtOH: Ethanol

FACS: Fluorescence activated cell sorting

FCS: Foetal calf serum

FISH: Fluorescent *in situ* hybridisation

FITC: Fluorescein isothiocyanate

FLIP: Fluorescence loss in photobleaching

FRAP: Fluorescence recovery after photobleaching

g: Gram

***g*:** Gravity

GTE: Glucose, Tris, EDTA

HDAC: Histone deacetylase

HU: Hydroxyurea

ICR: Imprinting control region

kb: Kilobase

kg: Kilogram

KSHV: Kaposi's sarcoma related herpes virus

l: Litre

LB: Luria-Bertani

LCL: Lymphoblastoid cell line

LCR: Locus control region

M: Molar

MAA: 3:1 methanol:acetic acid

MAGE: Melanoma antigen encoding protein

Mb: Megabase

MECP: Methyl CpG binding protein

MEF: Mouse embryonic fibroblast

mg: Milligram

min: Minute

ml: Millilitre

mm: Millimetre

mM: Millimolar

mmol: Millimoles

Mol: Moles

mRNA: Messenger RNA

MW: Mann-Whitney

µg: Microgram

µl: Microlitre

µm: Micrometre

µM: Micromolar

µmol: Micromoles

NEB: New England Biolabs

ng: Nanogram

nl: Nanolitre

nm: Nanometre

nM: Nanomolar

nmol: Nanomoles

OMIM: Online Mendelian Inheritance in Man

ORC: Origin recognition complex

ORF: Open reading frame

PBS: Phosphate buffered saline

PcG: Polycomb group

Pcdhb: Protocadherin beta

PCR: Polymerase chain reaction

pFA: Paraformaldehyde

PI: Propidium iodide

Plk: Polo-like kinase

PP2A: Protein phosphatase 2A

PSCS: Precocious sister chromatid separation

qPCR: Quantitative PCR

r^2: Interphase separation squared

r^2/a: Interphase separation squared divided by nuclear area

RBS: Roberts Syndrome

rDNA: Ribosomal DNA

RITS: RNA-induced transcriptional silencing

RNA: Ribonucleic acid

RNAi: RNA interference

ROI: Region of interest

RPMI: Roswell Park Memorial Institute

Rpol: RNA polymerase

rRNA: Ribosomal RNA

RSC: Chromatin structure remodelling

RT-PCR: Reverse transcriptase PCR

RT+: Reverse transcription positive

RT-: Reverse transcription negative

s: Seconds

Scc: Sister chromatid cohesion

SDS: Sodium dodecyl sulphate

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

siRNA: Short interfering RNA

SMC: Structural maintenance of chromosomes

SSC: Saline sodium citrate

TAE: Tris acetate buffer

TBE: Tris borate buffer

TBS: Tris buffered saline

TE: Tris, EDTA

TEMED: *N*,*N*,*N'*,*N'*-tetramethyl-ethane-1,2-diamine

Tgm: tRNA gene-mediated

TR: Texas Red

tRNA: Transfer RNA

TrxG: Trithorax group

UV: Ultraviolet

V: Volts

V/v: Volume/volume

W/v: Weight/volume

Abstract

Cornelia de Lange Syndrome (CdLS) is a multisystem genetic disorder caused by mutations in the cohesin complex. It is believed that cohesin is able to regulate gene expression with CTCF by holding chromatin in topological complexes, such as active chromatin hubs, and that CdLS is caused by loss of these complexes causing aberrant gene expression.

In order to determine if loss of these complexes in CdLS resulted in a general change in the compaction of chromatin, I undertook a series of analyses of the nucleus in CdLS patient lymphoblastoid cell lines (LCLs), compared to wildtype, and later in RNAi knockdown models of CdLS. By fluorescent *in situ* hybridisation (FISH) I studied the chromatin compaction of different regions of the genome, and found that in some, but not all, CdLS cell lines, gene-rich regions have less compact chromatin compared to wildtype. RNAi knockdown of two proteins that are mutated in CdLS, NIPBL and SMC1, also resulted in decompaction of regions of the genome, however these were different regions than in the patient LCLs, perhaps due to variation between cell lines. This change was not due to the interaction between cohesin and CTCF, as I found that knockdown of CTCF did not result in changes in chromatin compaction. I have also looked at the published data for gene expression in CdLS, and in mouse and *Drosophila* models of CdLS, and have found no correlation between the genes misexpressed in CdLS in the three species, nor between three cell lines of the same species.

These data suggest that the variation in chromatin compaction observed in CdLS may not be due to an interaction between cohesin and CTCF, and that cohesin can act independently of CTCF to regulate gene expression.

1 Introduction

The cohesin complex holds sister chromatids together before separation at metaphase. Mutations in cohesin proteins or cohesin-associated proteins are known to cause Cornelia de Lange Syndrome (CdLS). It is now believed that cohesin is also able to regulate gene expression through tethering of promoter-enhancer interactions, through interaction with other proteins, and it is disruption of these interactions that causes misexpression of genes and the phenotypes of CdLS.

I speculated that alteration in the tethering of intra-chromosome interactions would result in altered organisation of DNA in the nucleus, and altered chromatin compaction. My aim, in this thesis, was to examine the ways in which chromatin compaction was affected by CdLS, and to consider the molecular causes of any such changes.

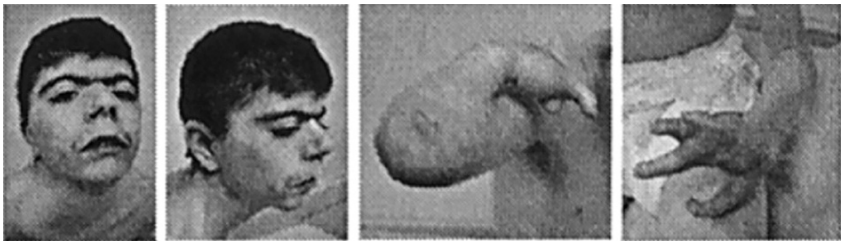
1.1 The cohesinopathies

1.1.1 Symptoms of the cohesinopathies

Cornelia de Lange Syndrome (OMIM 122470) is a genetic disorder that affects around 1 in 10,000 people, and results in a variety of developmental defects, including growth retardation, gastrointestinal problems, microcephaly, mental retardation, autistic-like behavioural problems, characteristic faces and hirsuitism, oro-dental problems, upper limb abnormalities, ophthalmic problems, seizures and cardiac problems (Figure 1.1A and B). There is a wide spectrum of severity in CdLS. The limb abnormalities range from severe deformities including monodactyly and complete absence of the radii (Figure 1.1A), to minor deformities such as small hands (Figure 1.1B). Variation can also be seen in the mental retardation phenotype of patients, as some patients are completely non-verbal, whilst others are capable of independent living.

Many similarities can be seen between CdLS and Roberts Syndrome (RBS; OMIM 268300). Similarly to CdLS, patients with RBS suffer from deformities of

A



B

C

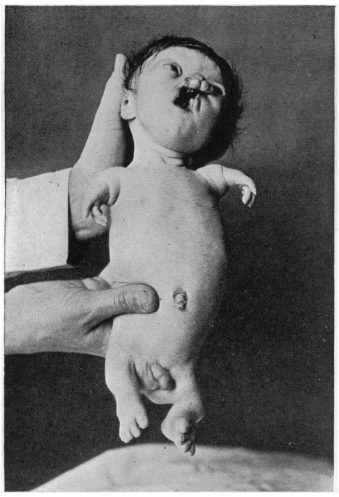


Figure 1.1 – Cohesinopathy patients.
A, a patient with mild CdLS (Gillis *et al* 2004). **B**, a patient with severe CdLS (Gillis *et al* 2004). **C**, a patient with RBS (from wikicommons).

the limbs (however in RBS these defects occur in both the upper and lower limbs), cleft palates, microcephaly and growth retardation (Figure 1.1C). Unlike CdLS, cells from RBS patients display altered metaphase chromosomes, with splaying out of the pairs of sister chromatids. RBS is generally more severe than CdLS.

In both syndromes, patients may show combinations of these defects at a whole spectrum of severity from classical or severe to mild. These diverse phenotypes are indicative of a developmental disorder in which the underlying molecular aetiology is aberrant gene expression in a variety of tissues.

1.1.2 Genetics of CdLS

CdLS mutations are mostly sporadic, with cases occurring in families with no known history. However there are families of multiple affected siblings, presumably due to parental germline mosaicism. Milder forms can be inherited from affected parents, but severely affected individuals will not have children, due to physical and mental disability, so these forms are never inherited (OMIM 122470).

CdLS was thought to be due to autosomal dominant mutations, which occurred spontaneously in most cases. Studies found that *NIPBL* was mutated in around half of all cases of CdLS (Krantz *et al*, 2004; Tonkin *et al*, 2004). Tonkin *et al* (2004) studied the breakpoints in previously identified CdLS patients with *de novo* translocations and compared them to other kinds of chromosome aberrations in these regions in other CdLS patients. A translocation between 5p13.1 and 13q12.1 in one CdLS patient corresponded to a deletion in another patient at 5p13.1-5p14.2. They mapped the breakpoint on chromosome 5 using fluorescent *in situ* hybridisation (FISH) and found that it mapped to a previously uncharacterised gene, which they named *NIPBL*, or *Nipped-B-like*, due to its sequence similarity to the fly *Nipped-B* (Figure 1.2). Krantz *et al* (2004) used a genome-wide linkage analysis and identified four candidate regions for a gene mutated in CdLS. The highest lod score, of 2.7, was also found at 5p13.1. Mutational analysis by conformation-sensitive electrophoresis identified *NIPBL*.

NIPBL is a large gene, and comprises 249,561 bp in 47 exons (Figure 1.3). The gene encodes a 2804 aa protein with a mass of 316 kDa. The NIPBL protein

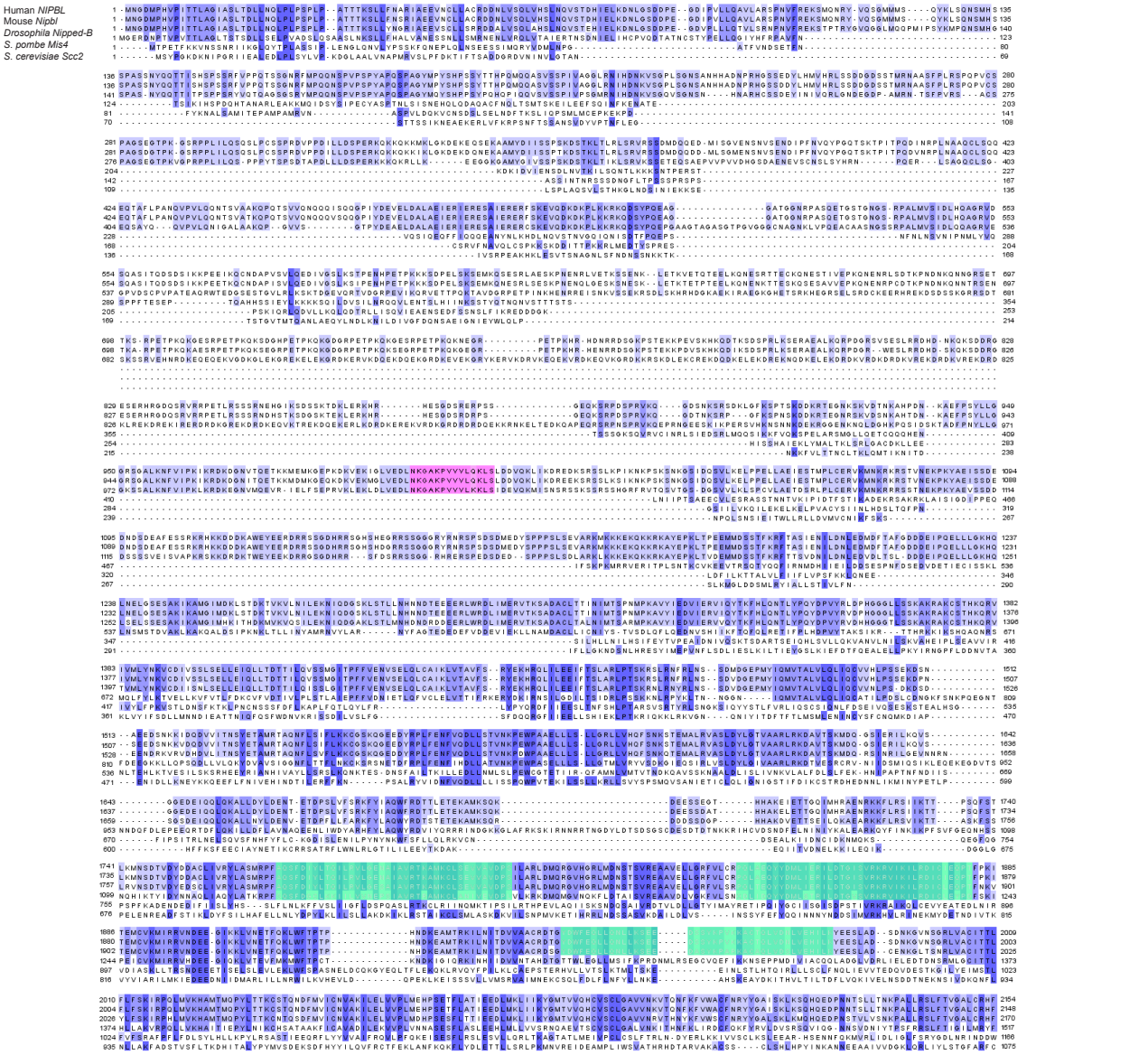


Figure 1.2 – NIPBL sequence alignment.
The sequence of NIPBL protein in humans (top) compared to the orthologous proteins in (from second from top to the bottom) mouse, zebrafish, *Drosophila*, *S. pombe* and *S. cerevisiae*. Bases are coloured blue to indicate identity between the species, with darker colour indicating higher identity. The PxVxL domain is highlighted in pink and the HEAT repeats are highlighted in green.

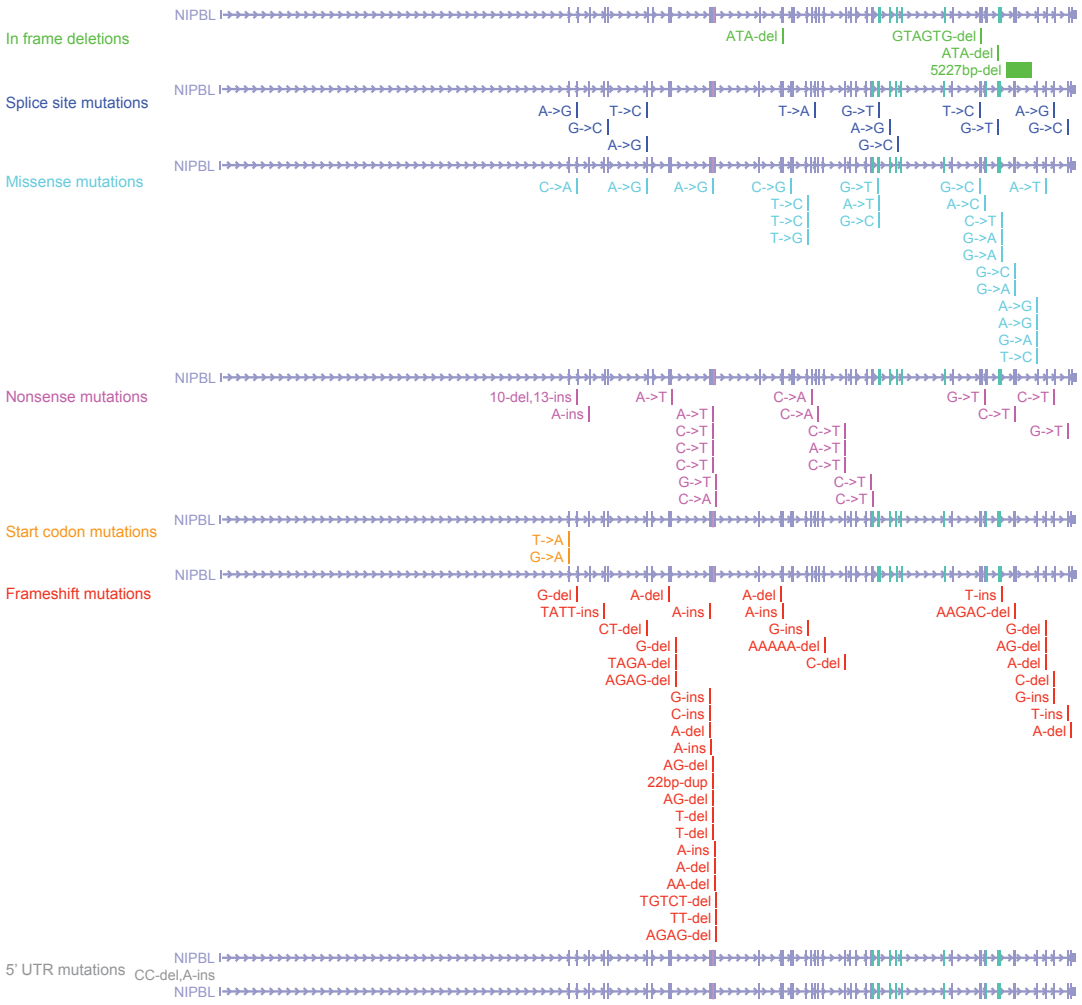


Figure 1.3 – CdLS mutations in NIPBL.
Track showing a number of NIPBL mutations identified in CdLS patients, created using the UCSC genome browser (GRCh37/h19 assembly; Kent et al, 2002; http://genome.ucsc.edu/). Mutations occur across the whole gene, and represent all possible mutation types. Exon 10, which encodes the PxVxL domain is highlighted in pink, and exons 27-33 and 38-40, which contain the HEAT repeats are highlighted in green.

contains a PxVxL domain that interacts with chromoshadow domains (Lechner *et al*, 2005), and glutamine-rich HEAT repeats near the C-terminus, which are thought to act as a scaffold for other proteins to assemble (Neuwald and Hirano, 2000).

NIPBL mutations were then identified by sequencing the gene in a variety of patients, and they included insertions, deletions, splice-site mutations, missense, nonsense and altered start codons (Figure 1.3), across the whole of the gene, with symptoms varying from mild to classical. All of these patients were heterozygotes, consistent with the dominant inheritance of the disorder, and this wide range of genotypes is indicative of haploinsufficiency. Patient cohorts have been studied to determine if there is a correlation between the type of *NIPBL* mutation and the severity of the clinical phenotype. Patient phenotypes were categorised as mild, moderate or severe, whilst genotypes were categorised as truncating, splice site or missense. A statistical correlation was found between patients with truncating mutations and severe phenotypes, whilst patients classified as mild or moderate were more likely to have missense or splice site mutations (Gillis *et al*, 2004; Bhuiyan *et al*, 2006; Selicorni *et al*, 2007).

NIPBL is mutated in only around half of all CdLS cases and mutations in other genes have since been implicated. Musio *et al* (2006) identified cases of CdLS with in-frame deletions and missense mutations in *SMC1L1* (OMIM 300590). Deardorff *et al* (2007) identified a single patient with a mild form of CdLS with an in-frame deletion in *SMC3* (OMIM 610759) and Zhang *et al* (2009) have identified a family with multiple affected siblings, all with a missense mutation in *PDS5B*. Both *SMC1L1* and *SMC3* mutations have only been reported in mild cases of CdLS. As these are all mild missense or in-frame mutations, this may indicate that more severe mutations of these genes are lethal. In the family affected by *PDS5B* mutations, a missense mutation was identified in two affected siblings, and the unaffected father and one unaffected sibling. The two affected siblings inherited a different maternal allele than their unaffected sibling, suggesting that the mother also carries a recessive allele that was not picked up in the screen and was passed on to her two affected offspring.

SMC1L1 is located on the X-chromosome but escapes X-inactivation and its effects are dominant. There are instances of maternal inheritance of *SMC1L1* mutation since some affected females can have children, possibly due to ratio of normal-to-mutated protein giving them only a mild phenotype. Males with *SMC1L1* mutation survive, however there are not sufficient numbers of *SMC1L1* mutants to statistically compare its incidence between the sexes. *SMC3* mutation is heterozygous in the one known affected individual, so its inheritance is assumed to be dominant. *SMC3* is found on chromosome 10. *PDS5B* mutation shows recessive inheritance and is on chromosome 13.

A large percentage of CdLS patients do not yet have a known gene mutation. *NIPBL* is implicated in around half of all cases, *SMC1L1* in less than 5%, *SMC3* mutation has only been reported once and *PDS5B* in only one family. It is clear that there is still much research to be done to find the rest of the genes that can be mutated to cause CdLS. Advent of whole exome sequencing should rapidly reveal this/these missing genetic component(s) (reviewed in Ng *et al*, 2010).

Complete heterozygous deletion of *NIPBL* appears to give a different phenotype to CdLS. A patient was born with a deletion at 5p13.2, encompassing the whole of *NIPBL* and no other known genes, but was not identified as a CdLS patient due to a complete lack of the facial phenotypes associated with CdLS. However, similarly to CdLS, this patient suffers from retardation of growth and development, limb defects, cardiac defects, cleft lip and palate and gastrointestinal reflux (Reeves *et al*, in preparation). *NIPBL* or *SMC1A* duplication gives a different phenotype to both CdLS and *NIPBL* deletion. Some features appear to completely contrast CdLS, including long philtrums and large hands with long fingers; other features are similar to CdLS, such as high palates, low set ears, low hairlines, mental retardation, developmental delay and sleep disturbances; however most features of CdLS are absent and it is clear that this is a distinct disorder (Yan *et al*, 2009).

1.1.3 Genetics of RBS

RBS (OMIM 268300) is caused by a mutation in *establishment of cohesion 1 homologue 2* (*ESCO2*). Unlike CdLS, RBS is autosomal recessive. Linkage analysis

in four families who shared a common ancestor in isolated villages near Bogotá, Colombia identified 8p12-21.1 as a locus for RBS. This interval was then narrowed down using more families and fine mapping, identifying a transcript which was later determined to be *ESCO2*. Mutations have been found in *ESCO2* in all studied RBS patients, and include frameshift, nonsense and missense mutations.

1.2 The cohesin complex

All of the genes currently known to be mutated in CdLS and RBS are involved in the function of the so-called cohesin complex, implicated in mitosis. SMC1L1 and SMC3 are subunits of the complex itself, NIPBL is the complex loader, and ESCO2 and PDS5 are required for the establishment of sister chromatid cohesion.

1.2.1 Subunits of cohesin are a members of the SMC family

The SMC family is a family of protein complexes consisting of cohesin, condensin and SMC5/6 in eukaryotes and bacterial bsSMC in *Bacillus subtilis*, RecN in *Deinococcus radiodurans* and MukBEF in *Escherichia coli*, with similar proteins in other bacterial species (Figure 1.4). In eukaryotes, these complexes are made up of a heterodimer of two structural maintenance of chromosomes (SMC) proteins; in cohesin these are SMC1L1 and SMC3, in condensin SMC2 and SMC4 and SMC5/6 is made up of SMC5 and SMC6. However *B. subtilis*, *D. radiodurans* and *E. coli* have only one SMC protein in each species, BsSMC, RecN and MukB respectively, all of which form homodimers. SMC proteins have anti-parallel coiled-coil domains and are around 50nm long, with globular head and hinge domains at either end of the coils. In the heterodimers of SMC proteins, the dimerisation interface of the two coiled-coils is the hinge domain. All SMC proteins have both their N and C termini at their head domains, and different proteins bind at this domain to complete the proposed ring structure of each complex (Losada and Hirano, 2005; Strunnikov, 2006; Lim and Oh, 2009). The kleisin protein family is closely associated with the SMC protein family, as they have been identified as closing protein rings (Schleiffer *et al*, 2003).

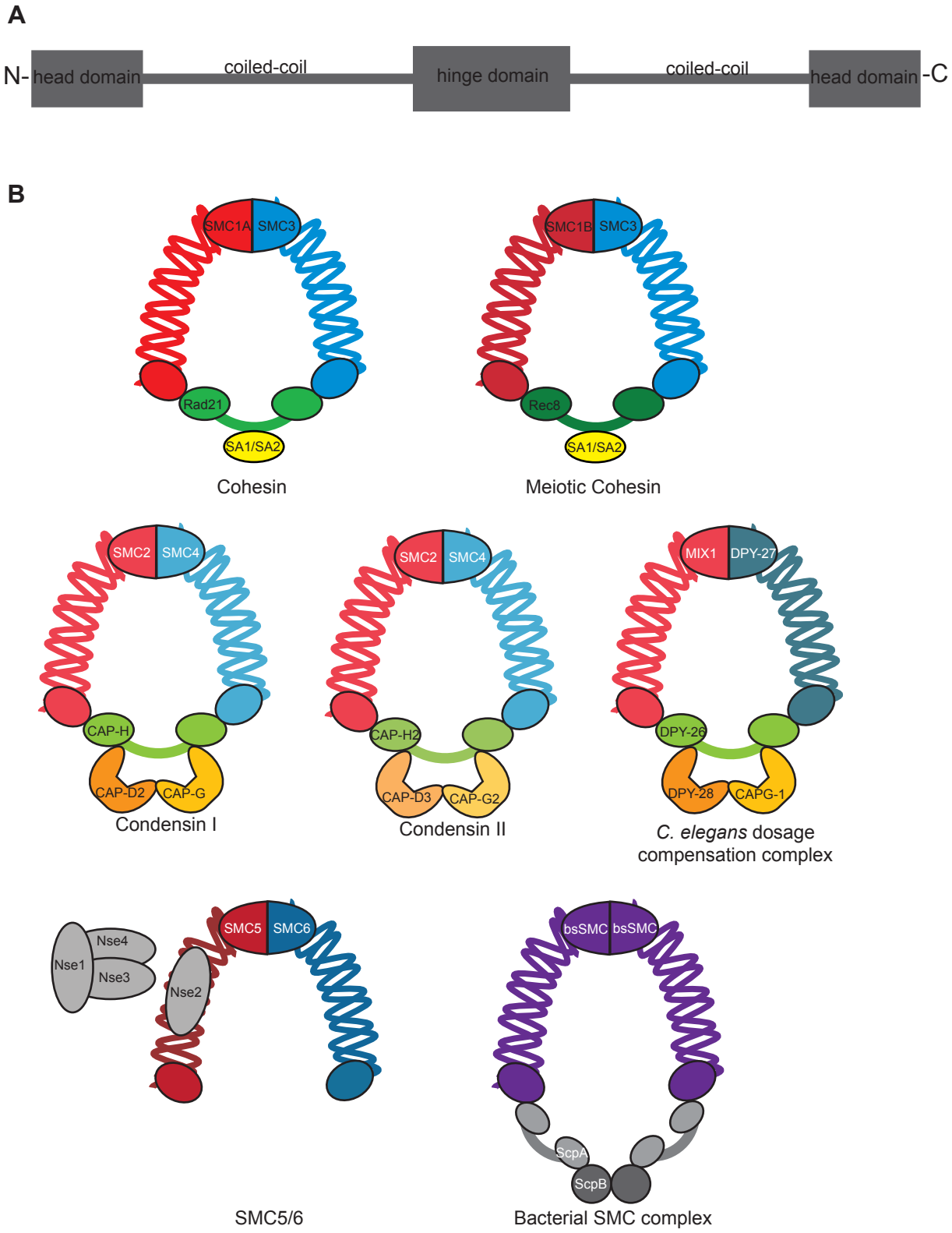


Figure 1.4 – SMC proteins and complexes.
A, scematic of the domains of SMC proteins. **B**, cartoons of SMC complexes. Adapted from Losada and Hirano (2005) and Wood *et al* (2010).

1.2.1.1 Cohesin

The cohesin complex is made up of four subunits; two SMC proteins, SMC1L1 and SMC3, and three kleisin proteins, Rad21 and SA1 or SA2 (stromal antigen; for ease I will always use the name of the human orthologue unless otherwise stated; Table 1.1; Figure 1.5). The complex is composed of two SMC protein arms, joined at the hinge region, and secured together at the head region by kleisin proteins. In cohesin, the gap between the SMC heads is bridged by Rad21. Either SA1 or SA2 binds to Rad21; all current evidence suggests functional homology of the two proteins. The resulting tripartite ring is 35nm in diameter. In meiotic cohesin, SMC1L1 is replaced by SMC1B and Rad21 is replaced by Rec8. The cohesin complex is loaded onto chromatin via a complex of NIPBL (Nipped-B in *Drosophila melanogaster*, Mis4 in *Schizosaccharomyces pombe* and Scc2 in *Saccharomyces cerevisiae*) and MAU-2 (Scc4 in yeasts; Figure 1.5; Losada and Hirano, 2005; Nasmyth, 2005).

1.2.1.2 Condensin

The protein complex condensin is related to cohesin and is involved in the compaction of chromosomes into their mitotic structure. Like cohesin, condensin consists of two SMC proteins, SMC2 and SMC4, forming a head-coiled-coil-hinge structure. A ring is formed by kleisin proteins, the chromosome associated proteins, CAP-G, CAP-H and CAP-D2 in condensin I and CAP-G2, CAP-H2 and CAP-D3 in condensin II, associating with the hinge domains of the SMC proteins (Figure 1.4). Compaction is speculated to occur by the formation of loops along the chromosome, held within the ring (Hagstrom and Meyer, 2003). Condensin II associates with the chromosomes in prophase and partially condenses the chromosomes by hierarchical folding, and is followed by condensin I in prometaphase, which then fully condenses the chromosomes by forming an axial glue structure. The two condensins appear to alternate along the axis of the metaphase chromatids (Losada and Hirano, 2005).

The condensin complex was first discovered and named in *Xenopus laevis*, as XCAP-C and XCAP-E (orthologues of SMC4 and SMC2 respectively) were found to form a heterodimer and complex with three other proteins, including an orthologue

Table 1.1 - Names of cohesin and cohesin-related genes in different species

Species	Human	Mouse	Zebrafish	Fruit-fly	Nematode worm	Budding yeast	Fission yeast
	Homo Sapiens	*Mus musculus*	*Danio rerio*	*Drosophila melanogaster*	*Caenorhabditis elegans*	*Saccharomyces cerevisiae*	*Schizosaccharomyces pombe*
Genes	*NIPBL*	*Nipbl*	*nipbl*	*Nipped-B*	*PQN-85*	*Scc2*	*mis4*
	SMC1L1	*Smc1L1*	*smc1a*	*SMC1*	*SMC-1*	*SMC1*	*psm1*
	SMC3	*SMC3*	*smc3*	*Cap*	*SMC-3*	*SMC3*	*psm3*
	Rad21	*Rad21*	*rad21*	*Rad21*	*SCC-1/COH-2* [a]	*Scc1*	*rad21*
	SA1 and SA2	*SA1 and SA2*	*SA1 and SA2*	*scc3*	*SCC-3*	*scc3*	*psc3*
	PDS5A and PDS5B	*Pds5A and Pds5B*	*pds5A and pds5B*	*pds5*	*evl-14*	*PDS5*	*pds5*
	MAU-2	*Mau-2*		*CG4203*	*MAU-2*	*Scc4*	*ssl3*
	ESCO1 and ESCO2	*Esco1 and Esco2*		*san and eco*	*F40F4.7*	*ECO1*	*eso1*
	Wapl	*Wapl*	*wapl*	*wapl*	*wapl-1*	*RAD61*	*wpl1*

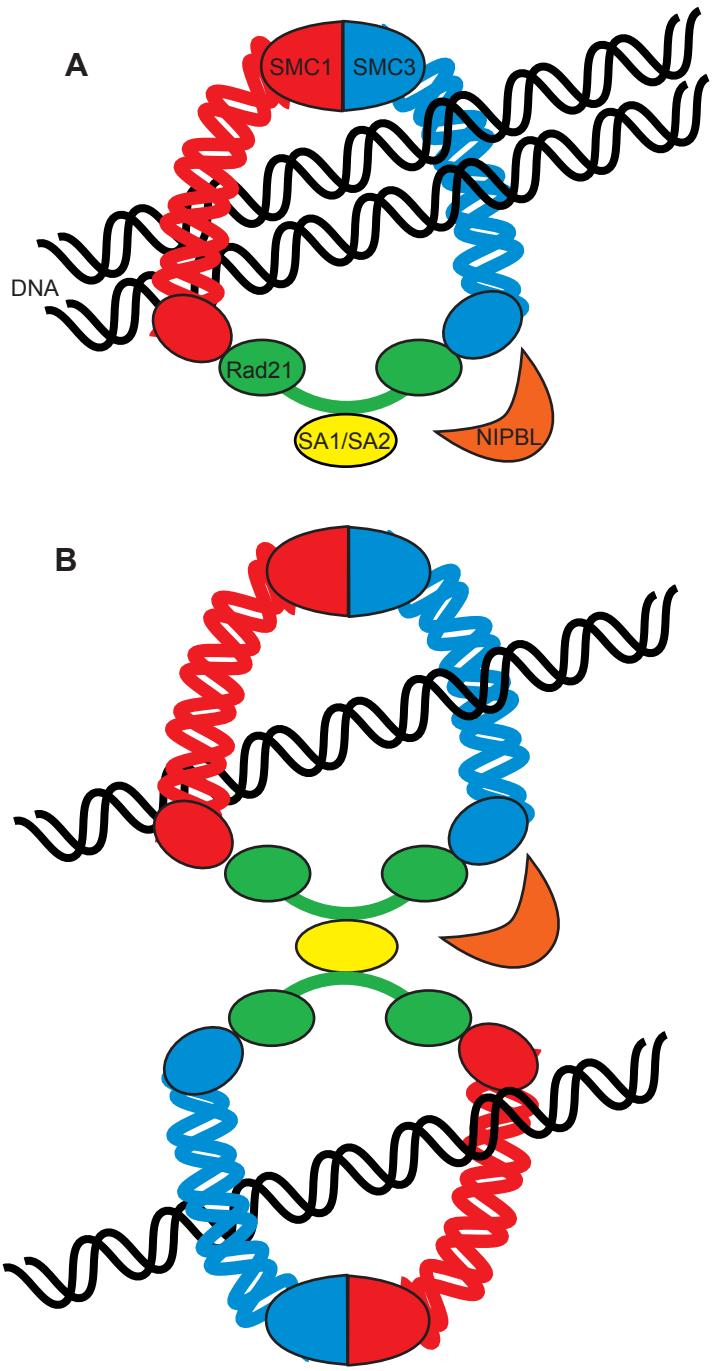


Figure 1.5 – The cohesin complex.
The human cohesin complex is made up of SMC1, SMC3, Scc1 and SA1/SA2. **A**, in the ring model a single cohesin ring encircles two strands of DNA (Haering *et al*, 2008). **B**, in the handcuff model, two rings each encircle a single strand of DNA and are held together by a single molecule of SA1 or SA2 (Zhang *et al*, 2008).

of the *Drosophila* Barren protein, which is now known to be an orthologue of CAP-H. The condensin complex was shown to be necessary for the condensation of chromatin at mitosis, independently of topoisomerase IIα (Hirano *et al*, 1997). Continuing to work with *X. laevis*, Kimura *et al* (1999) found that condensin is able to condense chromosomes by introducing positive supercoils, positive trefoil knots and an ordered global positive writhe. These actions require the presence of ATP, type I and type II topoisomerases and condensin phosphorylation.

In *Caenorhabditis elegans* and *Drosophila melanogaster*, condensin was also shown to be important in meiosis, promoting chromosome condensation after crossover and resolving linkages between sister chromatids. In flies, it is condensin II that is responsible for preventing transvection; similarly condensin II is responsible for this role in *C. elegans* too (Chan *et al*, 2004; Hartl *et al*, 2008).

The phenotypes of condensin mutants and knockdowns suggest various roles and actions of condensin. In *C. elegans*, depletion of SMC4 and MIX-1 (SMC-2 orthologue) results in aberrant chromosome structure and segregation at mitosis, and disrupted segregation at meiosis II but not meiosis I. However this depletion does not result in decondensation at metaphase, suggesting a complex role for condensin in chromosome architecture (Hagstrom *et al*, 2002). Similarly, the structural integrity of chicken chromosomes is lost upon conditional knockdown of SMC2, as is the localisation of non-histone proteins, such as topoisomerase II and INCENP in DT40 cells. In these cells, loss of SMC2 results in delayed, but eventually apparently normal but non-functional chromosome condensation (Hudson *et al*, 2003). In mouse ES cells, an RNAi screen found that knockdown of SMC2 and SMC4 results in metaphase arrest and enlarged interphase nuclei, without any change in DNA content (Fazzio and Panning, 2010).

Loss of condensin has also been shown to result in phenotypes that are unrelated to the cell cycle or chromatin condensation. In *Drosophila*, Barren (CAP-H) colocalises with Polycomb and is necessary for *Fab-7*-regulated homeotic gene expression (Lupo *et al*, 2001). Mice with a homozygous mutation in kleisin β (CAP-H2) have defects in T lymphocyte development, such that circulating T-cell numbers are decreased, there is increased expression of CD44 and levels of $CD4^+CD8^+$

thymocytes and their precursors are decreased. These mice are otherwise fertile, viable and have no cell cycle defects (Gosling *et al*, 2007). There is a strong enhancement of position effect variegation in *Drosophila* upon various condensin mutations, suggesting a role for condensins in interphase (Cobbe *et al*, 2006).

In *C. elegans*, a condensin homologue is important for dosage compensation of the X-chromosome in the hermaphrodite. The condensin in the DCC differs from normal *C. elegans* condensin by only one subunit, SMC4 is replaced with the specialised protein DPY-27. DCC is proposed to condense both X-chromosomes together into higher order structures, which reduce transcription from the two copies to a similar level as that in the XO males (Csankovszki *et al*, 2009).

Condensin-binding sites have been mapped in yeast and these colocalise with binding sites for the Scc2/Scc4 complex. Mutation of *scc2* or *scc4* decreased condensin-chromatin association and resultant chromosome condensation, indicating that Scc2/Scc4 promotes binding of condensin to chromatin, though it is not essential. A stable interaction between Scc2/Scc4 and condensin, however, could not be detected, so it is not impossible that Scc2/Scc4 is able to condense chromatin independently (D'Ambrosio *et al*, 2008).

1.2.1.3 The SMC 5/6 complex

The precise role of the SMC5/6 complex, the final member of the eukaryotic SMC family, is yet to be elucidated, though it appears to be involved in the DNA damage response, at double strand breaks, at recombination and replication forks and for the replication and stability of rDNA (De Piccoli *et al*, 2009). Other than the two SMC proteins, SMC5 and SMC6, the SMC5/6 complex also contain six non-SMC elements, Nse1-6, in yeasts. Nse1 has a RING finger motif and is suggested to be a ubiquitin ligase; Nse2 has a characteristic SUMO-ligase RING finger motif, which is able to sumolate some SMC5/6 complex subunits; Nse3 is a melanoma antigen encoding protein (MAGE), which are widely expressed in cancers (it is impossible to determine which of the 55 human MAGE proteins is orthologous to Nse3); the function of Nse4 is unknown; Nse5 and Nse6 form a heterodimer and interact with the hinge region of SMC5/6 (Figure 1.4; Duan *et al*, 2009). SMC5/6 is known to be

related to the DNA damage response and mutants of the complex are radiosensitive, and it has a role in the segregation of repetitive chromosome regions, such as rDNA repeat regions (Losada and Hirano, 2005). It is also required to dissolve DNA-mediated sister chromatid linkages prior to mitosis (Bermúdez-López *et al*, 2010). Scc2 is necessary for the localisation of the SMC5/6 complex on undamaged chromosomes, however in G_2/M the complex relocates to double strand breaks (DSBs) where SMC6 aids repair of collapsed replication forks; this relocation occurs independently of Scc2 (Lindroos *et al*, 2006).

1.2.1.4 Bacterial SMC complexes

BsSMC is not essential for survival of *B. subtilis*, however mutants have defects in chromosome segregation and condensation, suggesting the complex is functionally homologous to condensin. The bridge between the two head domains is made up of ScpA, similar to Rad21, and ScpB (Figure 1.4), which suggests that the complex is more structurally similar to cohesin. The structural similarity of the bacterial SMC complex to cohesin, but functional similarity to condensin suggests that both eukaryotic complexes may have evolved from duplications of the bacterial complex (Strunnikov, 2006).

Like BsSMC, the MukBEF complex in *E. coli* consists of a homodimer of an SMC protein, MukB, held in a ring by non-SMC subunits, MukE and MukF, which appear to be structurally homologous to ScpB and ScpA respectively. Unlike other SMC protein dimers, the MukB homodimer does not have a hole in its hinge domain. It is thought that MukBEF is able to condense the bacterial chromosome through the detachment of MukB head domains from MukF, entrapping DNA within the ring. MukBEF is found to from foci, colocalising with the *oriC* of the *E. coli* chromosome and forming distinctive clusters at the ¼ and ¾ positions in the cells (Lim and Oh, 2009).

RecN is a bacterial double strand break repair protein. Sequence analysis has identified it as related to the SMC proteins and biochemical analysis has found that homodimers of this protein bind duplex DNA, hydrolysing ATP. RecN is shown to stimulate intermolecular ligation of DNA molecules in the presence of DNA ligase,

presumably by holding the two ends of the DNA together. Little is known about its structure (Reyes *et al*, 2010).

1.2.2 The discovery of the cohesin complex

SMC1L1 was first identified in a screen in *S. cerevisiae*, which found that *smc1*$^{+/-}$ mutants had an elevated rate of minichromosome non-disjunction (Strunikov *et al*, 1993). Meiotic studies found that *smc1*$^{-}$ haploid spores did not survive. Linear chromosomes were then studied and non-disjunction, similar to that seen with the artificial minichromosomes, was again observed. Whilst *smc1* expression was observed throughout the cell cycle, budded *smc1* mutant cells showed clumping of the DNA at the bud neck and did not divide. Association of sister chromatids was one of the possible functions suggested for SMC1. The name "cohesins" was coined in a yeast screen for proteins involved in holding the sister chromatids together. Mutants were identified that lost chromosomes at high frequency during proliferation at 25˚C, and that could separate sister chromatids in the absence of APC/C function (the anaphase promoting complex) at 37˚C. The APC/C is a ubiquitin E3 ligase which marks numerous proteins for degradation to allow progression into mitosis, including cyclins and securin. Four complementation groups of mutants were identified in the screen; *smc1*, *smc3*, *scc1* (*sister chromatid cohesion*; the *Rad21* orthologue) and *scc2* (the *NIPBL* orthologue). Scc1 protein was most abundant and associated with chromosomes on entry into S-phase, but degraded during mitosis; this was dependant on the presence of Smc1 protein. The cohesins, SMC1, SMC3 and Scc1, were suggested to bind to sister chromatids as they replicate during S-phase before synchronous cleavage of the complex at mitosis allowing separation. The cleavage of the cohesin complex at anaphase was thought to be dependant on APC/C, as it was found to be necessary for dissociation of Scc1 from chromatin (Michaelis *et al*, 1997). Indeed, immunodepletion of SMC1 and SMC3 in *Xenopus laevis* results in normal nuclear envelope assembly and DNA replication but causes defects in the separation of sister chromatids in mitosis; these data reinforce the idea that these proteins have a role in cohesion (Losada *et al*, 1998). Synchronous separation of the chromosomes occurs as a result of cleavage of the Rad21 subunit of

cohesin by separase, a cysteine protease related to caspases (Uhlmann *et al*, 2000). Separase is inhibited throughout the cell cycle by Securin, however at anaphase Securin is ubiquitinated and targeted for degradation by the APC/C complex (Cohen-Fix *et al*, 1996). This releases Separase, which cleaves all the cohesin complexes associated with the kinetochores. Cohesin dissociates from the chromatids, the sisters are pulled to the opposite poles and the cell divides (Zou *et al*, 1999).

The role of the cohesin loader, NIPBL, was first identified by studying its *S. pombe* homologue, Mis4. Mis4 was initially identified as necessary for stable maintenance of minichromosomes in fission yeast, and mutations in this protein result in misegregation and premature separation of chromosomes. Missegregation was found to be due to a much earlier event than mitosis itself, since viability was lost during S-phase (Furuya *et al*, 2008). Later, the Mis4/NIPBL budding yeast homologue, Scc2, was found to complex with Scc4 (MAU-2 in humans) and mutation of either protein resulted in failure of cohesin complexes in binding DNA. Temperature sensitive mutants of *scc2* suffered from precocious sister chromatid separation (PSCS) at the restrictive temperature, and the resultant drop in cell viability occurred at the time of DNA replication. Loss of Scc2 at any later point in the cell cycle did not give a phenotype until the following cell cycle. Cohesin complexes were found in *scc2*$^{-/-}$ cells, but they were not able to associate with DNA, either at the centromeres or on the chromosome arms. This suggested that the protein was only involved in the establishment of cohesion and not in cohesion itself (Ciosk *et al*, 2000). Scc2/Mis4/NIPBL and Scc4/MAU-2 were dubbed “adherins” due to their role in cohesion, without being a member of the cohesin complex itself.

1.2.3 Models of sister chromatid cohesion by cohesin

Various models of cohesin structure and sister chromatid cohesion have been proposed (reviewed in Nasmyth and Haering, 2009). Most are variations on two basic models, one where the SMC1L1 and SMC3 head regions bind to the DNA in various conformations, and the other where the cohesin complex forms a ring encircling the DNA (Figure 1.5). Variations have been proposed for both of these basic models where cohesin acts as either a monomer or a dimer. Rings models have

long been favoured as these demonstrate how cleavage of the complex can result in both dissociation from the DNA and destruction of the complex.

Some *in vitro* evidence supports the encircling model (Figure 1.5A; Haering *et al*, 2008). Working on *S. cerevisiae* proteins, the group covalently linked the cohesin ring together; SMC3 and SMC1 were linked together at the hinge region by mutating adjacent side chains in the two proteins to cysteines, and the same was done at the head region of SMC1 and the C-terminus of Scc1. Upon incubation with dibromobimane (dBBr) or bis-malemidoethane, thiol linkages were formed between the cysteine residues of the proteins, holding the complex covalently together. A covalent link was made between SMC3 and Scc1 by creating a TEV-fusion protein, genetically fusing the two proteins together with a TEV cleavage site in between them; this link could then be cleaved using TEV protease, opening up the cohesin ring. The group had previously shown that cohesin complexes could dimerise circular minichromosomes (Ivanov and Nasmyth, 2007) and predicted that if cohesin held chromatids together by way of a ring structure, then minichromosome dimers held by a covalently closed ring would survive protein denaturation. This was found to indeed be the case when minichromosome dimers held with their covalently closed rings were treated with SDS, demonstrating that cohesin can encircle the DNA. These minichromosome dimers could be easily separated using TEV protease, showing that it is the ring structure that is necessary for dimerisation. To determine if these dimers were held by a single cohesin ring, or by dimers of cohesin rings, the efficiency of cross-linking was estimated at 30%. The efficiency of dimerisation was also found to be 30%, demonstrating that only a single ring of cohesin is needed to dimerise two minichromosomes. Evidence from Ghosh *et al* (2010) supports this model *in vivo* in the case of the *S. cerevisiae* 2 micron plasmid. This group used a fusion protein of Scc1 that contained a TEV cleavable element, and expressed this at an equal ratio to wildtype Scc1. Cleavage of Scc1 would then result in a 1:1 ratio of 2 micron plasmid dimers compared to separated dimers if the ring model were true, whilst any model requiring two cohesin rings would give a 1:3 ratio of dimers to separated dimers. The ratio of cleaved to uncleaved dimers was around 1:1, suggesting that pairs of DNA strands are enclosed by a single cohesin ring. Whilst this model is very convincing, we must consider the form that the chromatin takes in

this model; the cohesin ring is only 35nm in diameter, which would not be large enough to hold two 30nm chromatin fibres, so perhaps chromatin decondenses at regions of cohesin binding so that only the 10nm wide beads-on-a-string fibre interacts with cohesin.

In contrast, Zhang *et al* (2008) have evidence suggesting that cohesin forms a handcuff structure, made up of two cohesin rings, each holding a single DNA strand, held together by SA1 or SA2 at the two Rad21 linkers (Figure 1.5A). They show, by co-immunoprecipitation, a protein fragment assay and a yeast two-hybrid assay, that Rad21, SMC1L1 and SMC3 interact with themselves. SA1 and SA2 do not interact with themselves or each other, however one of them is necessary for the other cohesin subunits to interact with themselves. It is suggested that either SA1 or SA2, it does not matter which, holds the two cohesin rings together, holding the two Rad21 molecules in an antiparallel fashion. Inhibition of both SA1 and SA2 results in loss of Rad21-Rad21 interaction and loss of cohesion. This handcuff model would allow two 30nm fibres to be held together by one complex of two cohesin rings. They suggest that, at S-phase, single cohesin rings encircle the unreplicated DNA. On replication, cohesin rings can go to either DNA strand; pairs of cohesin rings on opposite strands are then tethered together by SA1 or SA2. The handcuff model is not compatible with the evidence shown by Haering *et al* (2008), as dimers of cohesin would not survive protein denaturation, unless they were looped within each other like links in a chain. Also Haering *et al* (2008) do not covalently link Scc3 (homologue to SA1/SA2) to their covalently closed cohesin complex, suggesting that Scc3 is not necessary for cohesion. However it is possible that these two configurations both exist, but function under different circumstances, for example it is possible that the ring shape encircles pairs of sister chromatids at mitosis, whilst the handcuff shape holds together DNA loops for gene expression, or vice versa.

1.2.4 The role of cohesin in the cell cycle

During S-phase and G_2, replicated sister chromatids remain tightly associated with one another. At metaphase, the chromatids line up and the microtubule spindles connect the kinetochores of the sisters to the spindle poles. The great tension on the

kinetochores as the spindles pull the chromatids from the opposite poles is necessary for the chromatids to line up perfectly, but is not sufficient to pull the two sisters apart until anaphase, when all the chromatids separate with flawless synchronicity. This synchronicity is achieved by cohesin (Nasmyth, 1999; Figure 1.6).

1.2.4.1 Cohesin loads onto the chromatin in G_1

Cohesin is loaded onto DNA during G_1, dependent upon Rad21 synthesis (Figure 1.6). Rad21 is synthesised in late G_1, before it is degraded at the metaphase-anaphase transition. Rad21 is able to associate with chromosomes if it is synthesised after S-phase, however it fails to establish cohesion (Uhlmann and Nasmyth, 1998). Cohesin rings form and load onto the single chromatin strand in late G_1. Both SMC proteins have an ATP-binding motif at both their N and C termini, which come into contact in the head domain. Mutagenesis of these binding motifs results in formation of the tripartite ring but this ring is not stable enough for cohesin function, which gives loss of viability due to lack of sister chromatid cohesion (Mishra *et al*, 2010). ATP binding is necessary for Rad21 association with both SMC1L1 and SMC3 but hydrolysis of ATP is not (Arumugam *et al*, 2003). Loading of the complex onto DNA requires NIPBL and ATP hydrolysis. Inactivation of the yeast NIPBL homologue, Scc2, results in a phenotype whereby the tripartite ring is able to form, but does not load onto chromosomes. It is suggested that Scc2 acts by promoting ATP hydrolysis in the head domains of the SMC proteins, and this is necessary for the loading of cohesin onto chromosomes (Ciosk *et al*, 2000). ATP hydrolysis is also the method by which NIPBL is thought to load condensin (D'Ambrosio *et al*, 2008; Lindroos *et al*, 2006). When the MAU-2 orthologue, Scc4, is depleted, mitotic chromosomes completely lack cohesin, and as a result arrest in prometaphase with PSCS (Watrin *et al*; 2006; Seitan *et al*, 2006). Fluorescence recovery after photobleaching (FRAP) studies in *Drosophila* suggest that Nipped-B converts cohesin into a stable, DNA encircling form, as depletion of Nipped-B reduces the amount of stable cohesin without reducing its residence time on DNA (Gause *et al*, 2010). In *Xenopus* egg extracts, the activity of NIPBL is shown to be dependent on replication licensing and the pre-replication complex, with loading of NIPBL

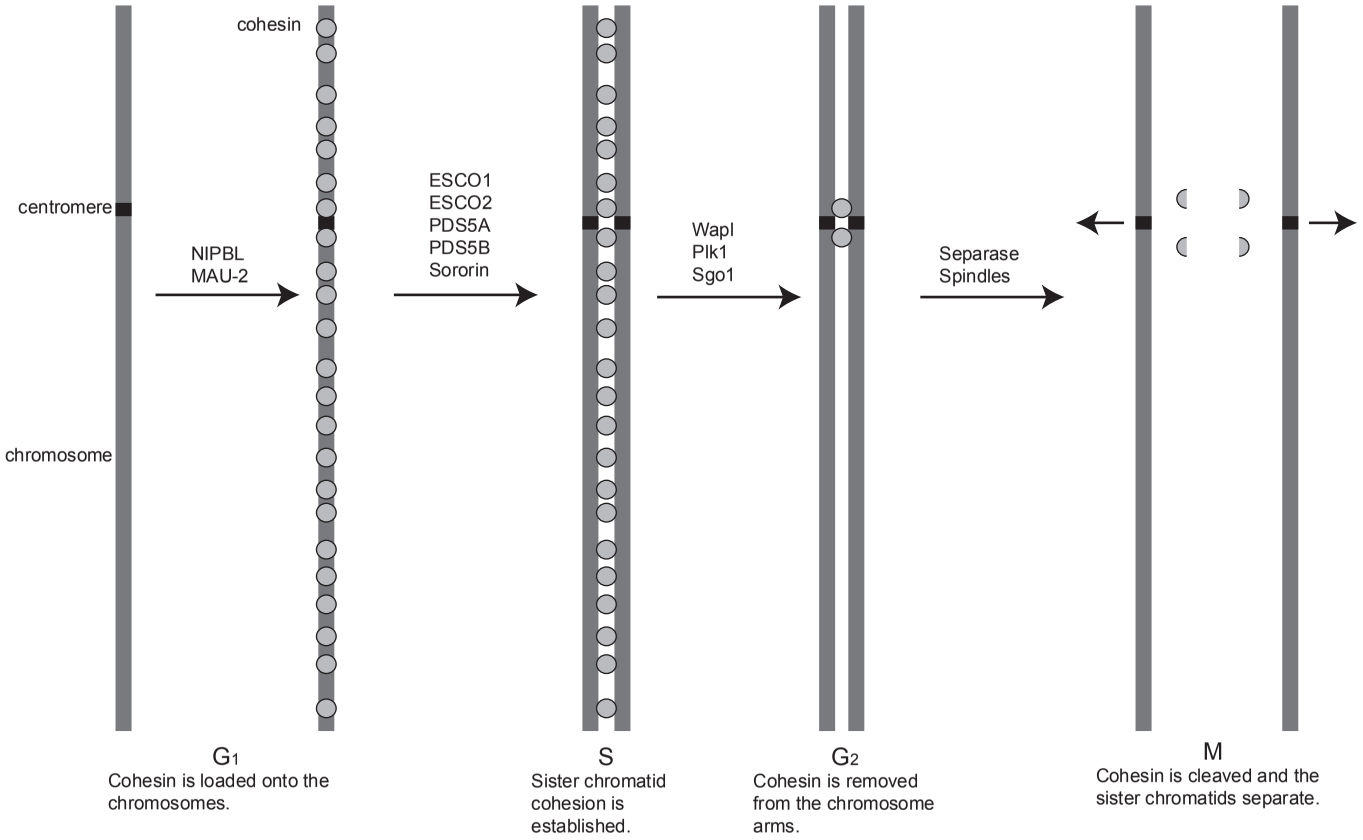


Figure 1.6 - Cartoon of cohesin activity through the cell cycle.

inhibited by geminin, a replication inhibitor that ensures only one round of DNA replication takes place (Gillespie and Hirano, 2004; Takahashi *et al*, 2004).

Cohesin binding at the centromeres is dependent on heterochromatin, but on the chromosome arms it is dependent on RSC (chromatin structure remodelling) in yeast. Loss of Swi6 (the *S. pombe* orthologue of heterochromatin protein 1 (HP1)) results in a complete loss of cohesin at the centromeres, however this does not alter localisation of cohesin on the chromosome arms (Bernard *et al*, 2001). Vertebrate NIPBL has a canonical PxVxL domain that is known to interact with HP1 proteins, and pull-down experiments show that this interaction does indeed take place in humans, and is abrogated upon mutation of the PxVxL motif (Lechner *et al*, 2005). However in human cells, RNAi of all three HP1 proteins (α,β and γ) does not affect bulk binding of cohesin to chromatin in interphase, or the pericentromeric accumulation of cohesin in metaphase (Serrano *et al*, 2009).

At the chromosome arms, the RSC nucleosome remodelling complex is found to be necessary for association with cohesin. In *S. cerevisiae* RSC cycles on and off the chromosome arms in a manner that precedes Scc1 (Rad21) binding. *Rsc* mutants exhibit cohesion defects, with chromosome arms coming apart prematurely but the centromeres remaining intact, as Scc1 is not able to associate with the chromosome arms, but is still able to associate with the centromeres (Huang *et al*, 2004). In human cells, Rad21 interacts directly with SNF2h, a member of the SWI/SNF nucleosome remodelling protein family, in a chromatin remodelling complex. This interaction is necessary for cohesin activity, as mutation of the ATPase activity of SNF2h results in disruption of Rad21 association with chromatin. 5-Aza-cytidine treatment of cells, which depletes DNA methyltransferases (DNMTs) thus decreasing the level of DNA methylation, increases Rad21 and SNF2h association with Alu repeats, suggesting that cohesin preferentially binds to unmethylated DNA (Liu *et al*, 2010). A correlation has also been observed between cohesin binding and acetylation of histones H3 and H4 (Hakimi *et al*, 2002).

After loading, the localisation of cohesin along the genome varies between species, as determined by ChIP. In yeasts, cohesin binds approximately every 10kb along the genome with narrow peaks of cohesin binding. Cohesin relocalises to sites

of convergent transcription and in *S. pombe* cohesin has been shown to promote transcription termination at these sites (Legronne *et al*, 2004; Gullerova and Proudfoot, 2008). However Kogut *et al* (2009) disagree and suggest that in budding yeast, cohesin is localised to sites of Scc2/Scc4 binding and not relocated by transcription. During G_1 Scc2/Scc4 has a poor affinity for cohesin binding sites, but as the cell approaches S phase, its affinity increases. Scc2/Scc4 then remains bound to the chromosomes and cohesin associates with it, and is loaded onto the chromatin.

In Drosophila, cohesin remains closely associated with Nipped-B. In contrast to yeasts, RNA polymerase is not able to relocate cohesin, and there is actually a strong correlation between the localisation of RNA polymerase and cohesin. Indeed cohesin binding varies depending on the expression status of a particular gene (Misulovin *et al*, 2008). Cohesin loading sites coincide with the origin recognition complex (ORC) in *Drosophila*, probably to allow establishment of cohesin immediately upon DNA replication (MacAlpine *et al*, 2010).

In mammals cohesin sites, determined by ChIP, colocalise with CTCF, the insulator protein (this will be discussed further in Section 1.5; Wendt *et al*, 2008; Stedman *et al*, 2008; Parelho *et al*, 2008). However other studies have found that cohesin colocalises with Mediator and tissue specific transcription factors, exclusively of CTCF (Section 1.6; Kernohan *et al*, 2010; Schmidt *et al*, 2010; Kagey *et al*, 2010).

1.2.4.2 Sister chromatid cohesion is established at S-phase

Replication of DNA in the presence of cohesin is not sufficient to establish cohesion. If Haering *et al* (2008)'s ring model is correct, then it is possible that as DNA replicates, DNA polymerase moves through the cohesin ring, which was loaded during G_1, and the two DNA strands remain inside the ring (Figure 1.5A). If the handcuff model proposed by Zhang *et al* (2008) is correct, then upon replication, cohesin bound to unreplicated DNA is shared between the two strands, then SA1 or SA2 is required to tether pairs of rings together (Figure 1.5B). Since we know that establishment of cohesion is an active process, I think it is likely that the second model is correct.

In yeasts, establishment of cohesion requires the acetyltransferase Eco/Ctf7, which acetylates two lysine residues in the ATPase head domain of Smc3 (Figure 1.6). Mutation of these lysine residues leads to cohesion defects and lethality in yeasts. In humans there are two homologues of Eco1, ESCO1 and ESCO2; both are necessary for cohesion. Although paralogous, the two proteins are not functionally homolgous, as only ESCO1 has been shown to acetylate SMC3 at the same residues as it is in yeast, and only ESCO2 has been implicated in RBS (Peters *et al*, 2008; Zhang *et al*, 2008). Mutation within the acetyltransferase domain has been shown to be sufficient to cause RBS (Gordillo *et al*, 2008). It is thought that Eco1 mediated acetylation is necessary to stabilise the interaction between cohesin and sister chromatids; indeed FRAP studies show that during G_1 cohesin interacts with chromatin in a dynamic manner, but during G_2 this interaction is stable (Gerlich *et al*, 2006). Acetylation of cohesin has also been shown to affect the speed of the replication fork, with replication progressing more quickly when cohesin is acetylated (Terret *et al*, 2009)

In *S. cerevisiae* Pds5, Scc3 (SA1/SA2 in humans) and Rad61 (the orthologue of Wapl) inhibit the establishment if cohesion until counteracted by Eco1 acetylation (Rowland *et al*, 2009), and in *S. pombe pds5* knockouts are able to establish cohesion in the absence of Eco1. In *Drosophila*, FRAP has shown that reduction of Pds5 increases the amount of cohesin that is stably bound to chromosomes (Gause *et al*, 2010). However Pds5 is also necessary to maintain cohesion during G_2, and there is synthetic lethality when it is mutated with either *mis4* or *Scc2* (Tanaka *et al*, 2001). In vertebrates there are two homologues of Pds5, PDS5A and PDS5B, which appear to be functionally redundant in sister chromatid cohesion (Peters *et al*, 2008). PDS5A and Wapl inhibit progression of the replication fork: in their absence DNA replicates quickly even if cohesin is unacetylated (Terret *et al*, 2009).

In vertebrates, Sororin is also necessary for cohesion; it binds to chromosomes upon the onset of DNA replication and is required both during and after S-phase, its knockdown leading to loss of sister chromatid cohesion. It associates with chromatin in prophase and a small amount remains on the chromosomes until metaphase, similar to the pattern of cohesin binding. Perhaps Sororin binding is also needed to stabilise cohesin's interaction with sister chromatids (Schmitz *et al*, 2007).

1.2.4.3 Cohesin dissociates from the chromosome arms at prophase

In vertebrates, much of the cohesin along the chromosome arms dissociates at prophase (Losada *et al*, 1998) and the cohesin left is bound only at centromeres; this removal is independent of Separase (Waizenegger *et al*, 2000; Figure 1.6). Fluorescence loss in photobleaching (FLIP) shows that in interphase the residence time of cohesin on chromatin is very short, ~60-90s, however that which associates with the centromere remains about 50% longer (McNairn and Gerton, 2009). Presumably, having cohesin bound only around a concentrated space better ensures synchronous cleavage of all the cohesin in the cell (Nasmyth and Schleiffer, 2004). Removal of cohesin from the chromosome arms is dependant on Wapl and Polo-like kinase 1 (Plk1) in vertebrates, which complex with Rad21 and SA1, and separately with PDS5A (Peters *et al*, 2008; Kueng *et al*, 2006). Loss of Wapl results in an increase in cohesin stably bound to chromosomes, shown by FRAP (Gause *et al*, 2010). Cohesin is removed by phosphorylation of SA2 by Plk1; although SA1 and Rad21 are also phosphorylated, this is not necessary for dissociation. Wapl is then required to dissociate cohesin from chromatin, perhaps controlling the ATPase activity of SMC1L1 and SMC3 such that cohesin rings are able to open. The centromere is protected from the activity of Plk1 and Wapl by Shugoshin (Sgo1), which is recruited to the centromeres by Bub1. It is thought to revert Plk1 phosphorylation of cohesin by recruiting protein phosphatase 2A (PP2A) to the centromere (Peters *et al*, 2008). In *Xenopus* egg extracts Wapl and Pds5 directly modulate changes in the conformation of cohesin, such that it can dissociate from chromatin during prophase, acting antagonistically with Sgo1 (Shintomi and Hirano, 2009).

1.2.4.4 Cohesin is cleaved by separase at metaphase

Synchronous separation of the chromosomes occurs as a result of cleavage of the Rad21 subunit of cohesin by separase. Separase is inhibited throughout the cell cycle by Securin, however at anaphase Securin is ubiquitinated and targeted for degradation by the APC/C complex. This releases Separase, which cleaves all the cohesin complexes associated with the kinetochores. Cohesin dissociates from the

chromatids, the sisters are pulled to the opposite poles and the cell divides (Zou *et al*, 1999). NIPBL dissociates from the chromatin upon entry into mitosis in a process that is dependant on Cdc2 in *Xenopus* (Gillespie and Hirano, 2004). SMC1 and SMC3 interact with the spindle pole bodies at mitosis. SMC1 also interacts with Rae1 (RNA export 1), and RNA binding protein that interacts with the cytoskeleton, and phosphorylation of SMC1 is necessary for this interaction. SMC1 is localised to the centre of the microtubule aster at anaphase, thought to be recruited by Rae1. Transfection with SMC1 and SMC3 resulted in formation of multipolar spindles, however a phenotype for depletion of either protein is not available (Wong and Blobel, 2008).

1.2.5 Cohesion in cohesinopathies

Based on cohesin's role in mitosis, one would predict that mutation in the complex, or proteins associated with it, would lead to aneuploidy giving a predisposition towards cancer in humans. Also cohesin is found to be necessary for the function of DNA-damage checkpoints in G_2/M and intra-S, by activating the checkpoint kinase, Chk2 and recruiting the checkpoint mediator 53BP1, independently of its cell cycle function (Watrin and Peters, 2009). Loss of this checkpoint would lead to increased DNA damage which could also contribute to carcinogeneis. However, there are no reports of increased infant or childhood cancers in CdLS, nor is cancer reported in adults with CdLS (Kline *et al*, 2007). In fact, the phenotype of CdLS is more indicative of development defects caused by altered gene expression. It has also been shown that all of the cohesin proteins are expressed in non-dividing differentiated cells that have no need for cohesin's cell cycle function (Wendt *et al*, 2008).

The *NIPBL* orthologue in fission yeast, *mis4*, was identified in screens looking for mutants defective in mitosis. Inactivation of *mis4* in G_1 resulted in cell death due to missegregation of chromosomes as a result of PSCS (Furuya *et al*, 1998). In *S. cerevisiae* mutation of *scc2* also causes PSCS, as cohesin complexes form normally but fail to associate with chromosomes (Ciosk *et al*, 2000). Homozygous mutations of *Nipped-B* in *Drosophila* also result in PSCS in 60% of the second-instar brain

cells studied, but there is not an increased level of aneuploidy in *Nipped-B* mutant cells (Rollins *et al*, 2004). However all of these studies used homozygous mutations, so cannot be extrapolated to CdLS, where mutations are heterozygous.

PSCS has also been reported in human CdLS cells, however its occurrence is extremely variable in both the controls in those experiments and CdLS cells and as such it is not considered to be a useful diagnostic tool for CdLS (Kaur *et al*, 2005; Castronovo *et al*, 2009). However in CdLS cell lines there is no evidence of aneuploidy (Pritchard, E., Master's Thesis 2007).

One might speculate that the intact checkpoints of these cells are able to initiate apoptosis as soon as PSCS is detected, resulting in the reduced growth rate observed both in CdLS patients and when culturing CdLS cells. This is consistent with work in zebrafish and human cells lines that demonstrates that knockdown of *SMC* results in apoptosis (Ghiselli, 2006). Attempts to generate mammalian cells deficient in SMC3 had resulted in apoptosis when SMC3 protein levels fell below a critical level. Morpholino-mediated knockdown of SMC3 in zebrafish resulted in necrosis at the five-somite stage, due to apoptosis, as seen with Acridine Orange staining. This was shown to be p53 dependant, as components of the p53 apoptosis pathway are upregulated in these embryos. Human cell lines lacking p53 were depleted of SMC3 by siRNA and were found not to enter apoptosis, whilst in p53 positive cells, the same siRNA treatment triggered significant apoptosis (Ghiselli, 2006).

CdLS mutations in the hinge domains of *SMC1L1* and *SMC3* affect the DNA binding properties of the two encoded proteins. Hinge dimers of these mutated proteins bind DNA with higher affinity than wildtype proteins. *SMC1L1* and *SMC3* mutated CdLS cell lines also display genomic instability (though no PSCS) and sensitivity to ionising radiation (Revenkova *et al*, 2009).

RBS cells show severe cohesion defects. There is PSCS, particularly at regions of repetitive gene sequences (German, 1979), often resulting in aneuploidy (Jabs *et al*, 1991). The splitting of sister chromatids in RBS lead them to be described as "railroad" chromosomes due to their resemblance to the two lines of railway tracks (Van den Berg and Francke, 1993). However there are no defects in homologous recombination or in sister chromatid exchanges in RBS (Van der Lelig *et al*, 2009).

Models of the cohesinopathies have been made in budding yeast, introducing mutations known in CdLS and RBS into the endogenous proteins, Scc2, Smc1 and Eco1. Unlike complete Scc2 and Smc1 knockouts, the CdLS mutations do not result in cohesion defects, similar to their human counterparts. The RBS mutations in Eco1 give railroad chromosomes, just as those seen in humans (Gard *et al*, 2009).

These data lead us to believe that cohesin has a role outside of sister chromatid cohesion. This idea is supported by evidence that in post-mitotic cells where there are no sister chromatids to cohese, mutation of cohesin still results in aberrant gene expression phenotypes (Pauli *et al*, 2008).

1.3 Cohesin is involved in gene expression

The clinical phenotypes of CdLS strongly suggest that cohesin has a role in gene expression. Indeed, evidence in many species suggests that this is the case, particularly by way of interaction with enhancers (Schaaf *et al*, 2009; Kawauchi *et al*, 2009; Liu *et al*, 2009; Skibbens *et al*, 2010; Kagey *et al*, 2010). Enhancers are regions of DNA, some distance away from a gene promoter, either in the 3' or 5' direction, that are able to interact with the promoter to give increased transcription. This mechanism by which enhancers work is not currently known, but they are speculated to either physically interact with the promoter, or proteins may be recruited to the enhancer then translocate along the DNA (reviewed in Sipos and Gyurkovics, 2005).

1.3.1 Yeast gene expression

Cohesin localises to sites of transcription termination in both budding and fission yeast. This localisation is not just a passive result of relocalisation by RNA polymerase, as previously hypothesised, but instead cohesin is specifically recruited to this location where it plays a functional role in preventing readthrough transcription (Legronne *et al*, 2004). Readthrough transcription often occurs where genes on opposite strands converge, resulting in the formation of dsRNA during G_1. It is found that these regions transiently take on heterochromatic features during G_1,

including trimethylation of H3K9 and association with Swi6, which also interacts with Scc3 and Scc2 (SA1/SA2 and NIPBL in humans). Formation of transient heterochromatin is dependant in the RNAi pathway, as deletion of the components of the Dicer or RITS (RNA-induced transcriptional silencing) complex resulted in loss of formation of this transient heterochromatin. In G_2, following recruitment to these regions by Swi6, cohesin then binds to these regions and transcription is terminated immediately following the end of the open reading frame (ORF). Cohesin does not localise to these regions unless both the converging genes are transcriptionally active (Gullerova and Proudfoot, 2008).

A role has been reported for Smc1 and Smc3 at boundary elements that separate heterochromatin and euchromatin in *S. cerevisiae* (Biswas *et al*, 2009; Donze *et al*, 1999). It was demonstrated that certain mutations of *smc1* and *smc3* abrogate the insulator activities of boundary elements flanking the repressed mating type *HMR* locus, resulting in the spreading of heterochromatin outside of *HMR* (Donze *et al*, 1999). Similarly at the ribosomal DNA locus, *RDN1*, the same *smc1* mutation resulted in loss of barrier activity. Interestingly, there is a known cohesin binding site, *CARL2* (cohesin associated region L2) in that region but the *RDN1* barrier does not coincide with it, rather occurring 2.4kb to the left of it (Biswas *et al*, 2009).

Genome-wide microarray expression data in *S. cerevisiae* suggests that even transient loss of Scc1 results in changes in transcription of a large number of genes (Skibbens *et al*, 2010). Many genes were up- or downregulated by culturing temperature sensitive *Scc1* mutants at the restrictive temperature, of which 29 were found to be reproducible, with expression changes greater than 1.5-fold. The genes affected appeared to be of related function, such as those involved in a single biosynthetic pathway, for cell wall metabolism or for 18S rRNA maturation. Also, misregulated genes were often in close proximity to one another, with 20% of the genes studied adjacent to another gene. However, the genes misregulated were not found to share any common motifs that might suggest where cohesin binds, nor did the group present ChIP data suggesting cohesin binding close to the misregulated genes.

1.3.2 Nipped-B, cohesin and regulation of gene expression in *Drosophila*

In animals, it appears that cohesin may be involved in the interactions between promoters and enhancers. The first evidence for this was found by Rollins *et al* (1999), who showed that mutation of *Nipped-B* increases the severity of *gypsy* insertion. Later evidence has also suggested that cohesin may be involved in axon pruning (Schuldiner *et al*, 2008; Pauli *et al*, 2008) and may also function as a chromatin remodeller (Hallson *et al*, 2008). The expression of many genes is altered in *Nipped-B* and *Rad21* knockdown cells of *Drosophila* (Schaaf *et al*, 2009).

Cut is an enhancer driven gene in *Drosophila*, expressed in the wing tip; *cut* mutations result in disruption of the wing tips, with phenotypes varying in severity from small gaps in the margin to loss of most of the bristles from the entire wing margin. Mutations within an enhancer region 80kb upstream of the *cut* promoter result in wing tip defects (Jack *et al*, 1991). *Gypsy* is a *Drosophila* retrotransposon that is able to act as an insulator – when it is inserted between an enhancer and promoter it is able to interfere with transcription from that promoter (Parkhurst and Corces, 1986). *Gypsy* insertion in between *cut* and its enhancer can result in wing nicks, such as are seen in either *cut* or *cut* enhancer mutation (Jack *et al*, 1991).

Nipped-B (the fly *NIPBL* orthologue) mutation has been shown to increase the effect of a *gypsy* insertion. The wing tips of *Nipped-B* mutants, which had a *gypsy* insertion in between *cut* and its enhancer, were more severely affected than those with a *gypsy* insertion but wildtype *Nipped-B*. This is certainly a synthetic phenotype, as *Nipped-B* mutants without the *gypsy* insertion did not have this same phenotype (Rollins *et al*, 1999). Mutation of *PDS5* gives a similar effect to *Nipped-B* on the *cut* wing phenotype (Dorsett *et al*, 2005). Heterozygous loss-of-function mutations in *Nipped-B*, similar to the *NIPBL* mutations found in CdLS, were able to dominantly increase the severity of various *cut* and *Ubx* mutations and reduce the expression of wildtype *cut* (Gause *et al*, 2008). In cultured cells, a cohesin binding site was found between a remote wing margin enhancer and the *cut* promoter. Reducing the expression of *Smc1* or *pds5* mutation in these cells, or knockdown Scc3 or Rad21 in the wing tip gave an increase in *cut* expression, whilst Nipped-B

knockdown inhibits *cut* expression (Rollins *et al*, 2004; Dorsett *et al*, 2005). These data prompted the hypothesis that Nipped-B may regulate binding of cohesin at this region, where cohesin blocks interactions between the promoter and enhancer (Rollins *et al*, 2004).

A role has also been found for cohesins in *Drosophila* neurons. Deletion of *smc1* by *piggyback* insertion in neuroblast clones results in a loss of axon pruning. At puparium formation in larval development, axons normally extend into the dorsal and medial lobes, but 18 hours later they are pruned back. In *smc1* mutants these axons are not pruned, and these unpruned axons persist into the adult brain. This phenotype can be rescued by post-mitotically expressing *smc1* in neurons (Schuldiner *et al*, 2008). TEV protease cleavage of Rad21 gives a similar phenotype with loss of axon pruning after puparium formation, persisting into the adult brain (Pauli *et al*, 2008). Similarly, MAU-2 is involved in axon migration guidance in *Caenorhabditis elegans* (Seitan *et al*, 2006; Watrin *et al*, 2006). It is clear that cohesin proteins are necessary for post-mitotic axon organisation in invertebrates, but whether this is due to altered long-range chromatin interaction is unknown.

However *Drosophila* cohesins may also be able to regulate gene expression through chromatin remodelling complexes (Hallson *et al*, 2008). The *Drosophila* Rad21 protein is encoded by the locus *verthandi* (*vtd*), mutations of which are associated with trithorax group (*trxG*) phenotypes. *TrxG* genes are implicated in transcriptional regulation, and can regulate homeotic genes, often acting antagonistically to polycomb group (*PcG*) genes which maintain the repressed state of homeotic genes. One of the genes regulated by *vtd* is *hedgehog* (*hh*), and *vtd* suppresses dominant gain-of-function alleles of *hh*. Like *vtd* mutations, *Nipped-B* mutations dominantly suppress *PcG* and gain of function *hh* mutations, however mutations in *Pds5* and *Smc1* had no effect on *PcG* or *hh*. *Vtd* displayed a role in sister chromatid cohesion as mutations were able to disrupt the cell cycle due to failure of initiation and maintenance of sister chromatid cohesion.

In flies, cohesin colocalises with active genes and RNA polymerase II, and away from PcG-mediated silencing marks (Misulovin *et al*, 2008). Genome wide expression microarray data shows that genes with altered expression after RNAi of

Nipped-B or Rad21 are localised close to cohesin binding sites, and that there is a linear correlation between genes affected by RNAi of Rad21 and those affected by RNAi of Nipped-B (Schaaf *et al*, 2009). There are some genes, such as the *invected-engrailed* complex, and the *Enhancer of split* complex (*E(spl)*), that appear to be hypersensitive to loss of either Nipped-B or cohesin. Weak knockdown of either Nipped-B or Rad21 gives a small decrease in expression of these genes, whilst strong knockdown gives up to a 130-fold increase in expression. These genes occur in regions that have both PcG and TrxG marks in pluripotent cells; it is thought that these genes may be maintained in this bivalent state, with control by cohesin, so that they can be switched on or off very quickly. Indeed this may suggest why Rad21 has been identified as a TrxG protein (Hallson *et al*, 2008). However *invected-engrailed* and *E(spl)* are by no means typical, the effect of cohesin knockdown on the expression of most genes is less than two-fold (Schaaf *et al*, 2009). Perhaps cohesin controls enhancer and promoter interactions at most locations, but also plays a role in maintenance of bivalent state at other regions.

1.3.3 The effect of cohesin and NIPBL on vertebrate gene expression

In the zebrafish *Danio rerio*, both cohesin and condensin genes, including *smc1a*, *rad21*, *pds5b*, *smc3*, *smc2* and *smc4*, are expressed in proliferating regions during development, overlapping with expression of *pcna*. However, cohesin genes are additionally expressed in non-proliferating regions in the mature zebrafish brain and eye, suggesting that as in invertebrates, cohesin has a role in zebrafish other than in sister chromatid cohesion (Mönnich *et al*, 2009).

Like in *Drosophila*, decreased cohesin function in zebrafish results in altered gene expression, such as that of runx transcription factors, which determine the fate of many cell lineages, with roles in both proliferation and differentiation. An *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen in zebrafish, searching for regulators of *runx1* expression by *in situ* hybridisation, identified a nonsense mutation in *Rad21*. *Runx1* expression was lost in the early embryo with the *Rad21* mutation, and embryos suffered from a developmental delay due to a block in mitosis. Other

haematopoietic transcription factors were expressed normally in the mutants, but markers of differentiated blood cells were lost at later stages. Runx1 expression was also lost following morpholino knockdown of *smc3*. To determine if this effect was directly due to loss of cohesin or loss of progression through the cell cycle, embryos were treated with two cell cycle inhibitors, hydroxyurea (HU) and aphidicolin, neither of which replicated the effect of loss of cohesin (Horsfield *et al*, 2007).

Morpholino knockdown of MAU-2 or NIPBL in *Xenopus tropicalis* results in phenotypes that are reminiscent of CdLS. Both knockdowns result in delayed development from gastrula stages, and at later stages the embryos exhibit defects in neural, somite, head, eye and tail development, with truncated anterior-posterior axes and ventralisation (Seitan *et al*, 2006; Watrin *et al*, 2006).

1.3.4 Cohesinopathy mutations affect gene expression in mammals

The first CdLS-like mouse model was generated by mutation of *pds5b* (Zhang *et al*, 2007; Zhang *et al*, 2009; Figure 1.7). It gives a phenotype very similar to that found in CdLS, including small size, short limbs, small head and facial dysmorphisms resembling those of CdLS patients. Alcian Blue staining revealed the upper limbs had shortened bones, but not the limb truncations or fused digits seen in severe cases of CdLS, however such major limb abnormalities were also not observed in the human family with *PDS5B* mutation (Zhang *et al*, 2009). The mice showed complete secondary palate cleft and non-meeting palatal bones. Atrial septal defects, ventricular septal defects and atrioventricular canal defects were observed in the mutant mice, as with CdLS patients. No obvious cohesion defects, such as PSCS, were observed in these mice.

A *NIPBL* knockout mouse model of CdLS has more recently been generated (Kawauchi *et al*, 2009; Figure 1.7). A gene-trap in the intron between exon 1 and exon 2 (this intron occurs prior to the transcription start site) gave complete loss of protein expression from that allele, however Nipbl protein levels were only reduced by 25-30%, similar to in humans with CdLS, suggesting that Nipbl may be autoregulatory. Heterozygotes showed 75% lethality at birth, with a third of the

remaining mice failing to thrive and dying in the first two weeks; surviving mice display "catch-up" growth but never achieve wildtype size. The adults show similar phenotypes to CdLS patients, including abnormal faces, cardiac defects, eye abnormalities, microcephaly, repetitive behaviours, seizures, auditory defects and shortened long-bones. No severe limb abnormalities, similar to those seen in CdLS, were observed, nor were cohesion defects observed. Microarray analyses were performed on RNA extracted from mouse embryonic fibroblasts (MEFs) generated from *NIPBL*$^{+/-}$ embryos, and from the brains of these mice. 978 genes in the brain extracts, and 81 genes in the MEFs were subtly misregulated in the CdLS models compared to wildtype. 97% of genes misexpressed showed less than a 1.5-fold change in expression, and less than 0.4% of genes had more than a two-fold change. However it is difficult to determine which genes are directly affected by the mutation, and which were knock-on effects of other gene expression changes. These data confirm the idea that CdLS is caused by subtle changes in expression of a large number of genes, rather than gross effects on a small number of genes. This is consistent with cohesin acting as a general regulator of gene expression, rather than as a specific transcription factor (Kawauchi *et al*, 2009).

Some of the misregulated genes identified in mice may give us clues to the phenotypes we observe in humans in CdLS, and in the mechanism by which cohesin can affect gene expression. Identified in the brain study were some genes that are known to be associated with autism and epilepsy, suggesting a possible mechanism for the mental retardation, behaviour and seizure phenotypes observed in CdLS patients. Studies in the MEFs identified genes involved in adipogenic differentiation, and the MEFs were shown to be unable to differentiate into adipocytes, which may suggest a mechanism for the lean-ness of both CdLS patients and *NIPBL*$^{+/-}$ mice. Decreases in expression of all 22 genes of the *protocadherin beta* (*Pcdhb*) cluster on mouse chromosome 18 suggest that effects of Nipbl on gene expression are position specific. The strongest effects occurred in the genes at the 3' end of the cluster, which are already more highly expressed than those at the 5' end. These data suggest that there is an enhancer at the 3' end of the cluster, which depends upon Nipbl (Kawauchi *et al*, 2009).

A

Wildtype *Nipbl* gene-trap heterozygote

B

Wildtype *Pds5b-/-*

Figure 1.7 - Mouse models of CdLS.
A, A newborn mouse heterozygous for a *Nipbl* gene trap, compared to a wildtype littermate (from Kawauchi *et al*, 2009).
B, A newborn mouse homozygous for a *Pds5b* mutation, compared to a widtype litter mate (from Zhang *et al*, 2007)

Knockdown of Nipbl and SMC1 by siRNA in mouse embryonic stem (ES) cells also gives subtle misregulation of a large number of genes (Kagey *et al*, 2010). Genes downregulated in ES cells include the pluripotency associated transcription factors that give ES cells their "stem-ness", and genes upregulated include developmental transcription factors, suggesting that loss of Nipbl and SMC1 can cause early differentiation of stem cells.

Expression microarray studies have been carried out on mRNA from *NIPBL* mutated patient lymphoblastoid cells lines to determine the genes misexpressed in CdLS (Liu *et al*, 2009). mRNA from 16 CdLS probands compared to 17 healthy controls identified between 339 and 1,501 nonredundant genes, depending on the stringency of scoring, that were consistently differentially expressed in CdLS. As with the mouse model (Kawauchi *et al*, 2009), the changes in expression level for all genes are low; the mean fold change in either direction is 1.42, and the maximum upregulation 4.61-fold and the maximum downregulation 3.38-fold. A ten-gene expression profile for CdLS was created which was then tested on 80 probands, of which 62 suffered from CdLS of varying severity, mutations in different genes and unknown mutations, and four suffered from other congential multisystem disorders, to determine if it could predict incidence of CdLS based on gene expression alone. This algorithm correctly identified all CdLS patients, excluded all healthy cohorts and three of the four cohorts with other genetic disorders, however it did classify RBS as CdLS, demonstrating how similar the expression profile of RBS is to CdLS. The degree of misexpression of genes reflected the severity of the patients' conditions, which severely affected individuals showing high levels of gene misexpression, and milder cases with less misexpression. The microarray data was compared to Chromatin Immunoprecipitation (ChIP) for Rad21, and it was found that cohesin is often found bound close to the 5' UTR and the transcription start site of genes misregulated in CdLS, often overlapping with CTCF sites. Indeed, most misexpressed genes had a cohesin binding site within 1kb of the transcription start site. In CdLS, binding of Rad21 at these sites was reduced, suggesting a causal link between cohesin binding and gene expression (Liu *et al*, 2009).

A genome wide methylation study in CdLS cell lines found that there is little alteration in DNA methylation patterns on autosomal chromosomes in CdLS

compared to wildtype, demonstrating that DNA methylation is not the mechanism by which cohesin modulates autosomal gene expression in humans. However there are changes in DNA methylation on the X-chromosomes in CdLS cell lines: in females the methylation of hypomethylated sites on the X-chromosome is increased however in males it is decreased (Liu *et al*, 2010).

Gene expression may also be affected by the interaction of NIPBL with histone deacetylase 1 (HDAC1) and HDAC3. NIPBL is able repress promoter activity of a reporter gene by recruiting HDACs to the promoter, while CdLS mutations in *NIPBL* reduce this repression, similar to the effect seen in chemical inhibition of HDACs (Jahnke *et al*, 2008).

1.4 Chromatin and higher order structures

One possible mechanism for the general low-level effect of CdLS on gene expression may be through nuclear organisation and chromatin looping mediated by cohesin.

1.4.1 Gene activation and chromatin organisation

Chromosomes occupy specific regions of the nucleus, called chromosome territories (CTs). The position of a chromosome in the nucleus is dependant on its size, with smaller chromosomes found towards the interior and larger ones towards the periphery, and on its gene density, with gene-rich chromosomes towards the interior and gene poor ones at the periphery. Gene-rich regions and active genes tend to be located on the surface of the CT, whilst heterochromatic, gene-poor and inactive regions will be located within the CT (Boyle *et* al, 2001; Cremer and Cremer, 2001; Lieberman-Aiden *et al*, 2009; Mahy *et al*, 2002a).

Whilst the transcriptional machinery is able to access all chromatin in the nucleus, including that buried in the middle of CTs, this machinery is often concentrated within distinct foci, called transcription factories. Transcription factories contain aggregates of Rpol II and remain intact, even in the absence of transcription (Mitchell and Fraser, 2008; Sutherland and Bickmore, 2009). Some

genes actively loop out of their CTs upon activation. This looping has been demonstrated by FISH to occur at active gene-rich regions on chromosome 11 (Mahy *et al*, 2002b), across the HoxB locus during development (Chambeyron and Bickmore, 2004), at HoxD upon cell differentiation (Morey *et al*, 2007) and at the major histocompatibility complex upon activation (Volpi *et al*, 2000). It is through this looping that active chromatin hubs (ACHs) may form that become associated with transcription factories (Mitchell and Fraser, 2008), however looping out itself is not sufficient for gene activation, and transcription factories are able to associate with genes within the CT (Morey *et al*, 2009; Heard and Bickmore, 2007).

1.4.2 Chromatin folding follows a random-walk giant-loop model

At a higher order, chromatin is thought to follow a random-walk giant-loop model (Figure 1.8). Chromatin may be tethered together at intervals, forming "giant loops" but within those loops it follows a random walk. A random walk model assumes that chromatin behaves as a freely jointed polymer chain. This model was developed using data from FISH experiments. Pairs of fluorescently labelled probes with separations from 100kb to 4Mb were hybridised to the nuclei of fixed G_0/G_1 cells then imaged. The physical distance between the two probes was measured and compared to the known genomic distance between them. It was found that with separations between 100kb and 1.5Mb, the relationship between mean-squared physical distance and genomic distance was linear and conformed to a random walk model (Figure 1.8B; van den Engh *et al*, 1992; Mateos-Langerak *et al*, 2009).

Under a random walk model of polymer folding, the distance along the polymer will be directly proportional to the mean of the physical distance squared, taking into account greater variances as the distances get larger. A set of physical distance measurements of one known polymer distance will conform to a Rayleigh distribution. A Rayleigh distribution is similar to a Normal distribution, however the variable can extend to infinity (or a sufficiently high number that it appears to extend to infinity) to the right of the distribution, and has a discrete cut-off, often zero, to the left of the distribution (Figure 1.8C). The result of this is that the distribution of data is skewed to the right. A characteristic of a Rayleigh distribution is that the mean of

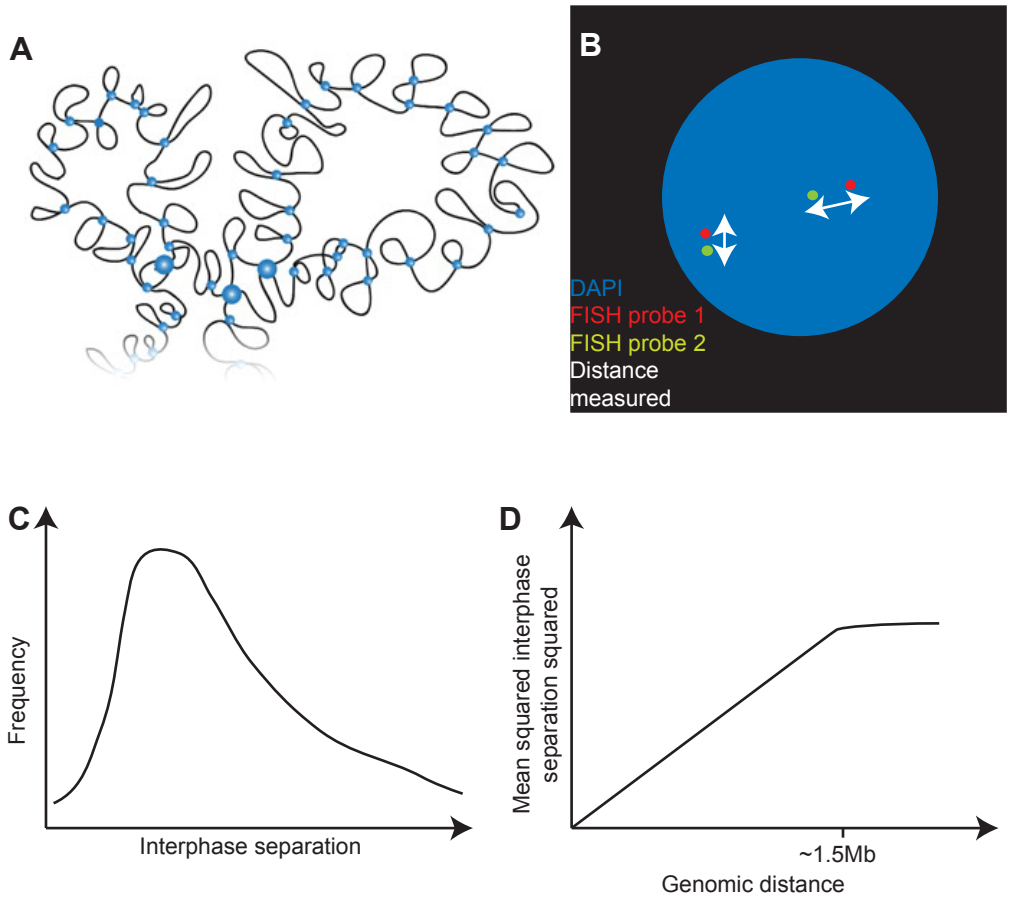


Figure 1.8 – The random-walk giant-loop model of chromatin compaction.
A, an illustration of the random-walk giant-loop model; in this model of polymer folding, the polymer follows a random walk up to a certain distance, then is tethered together. In the case of chromatin this is thought to be via proteins (blue circles) every ~1.5Mb (from Mateos-Langerak *et al*, 2009). **B**, an illustration of a FISH probed nucleus testing the random-walk giant-loop model. Probes of known distance apart are hybridised to interphase nuclei and the physical distance between them is measured. **C**, an illustration of a Rayleigh distribution obtained from a FISH experiment. **D**, an illustration of the relationship between genomic distance and mean interphase distance squared. These values are directly proportional up to the genomic distance of ~1.5Mb (van den Engh *et al*, 1992).

the data will always be slightly higher than the median (median: mean ~ 0.9) and the standard deviation will always be around half the value of the mean (van den Engh *et al*, 1992).

Above 1.5Mb, there must be some other chromatin organisation happening. A giant-loop model is the simplest way to model this, and it is predicted that the giant loops in chromatin are around 2-3Mb and secured by proteins (Sachs *et al*, 1995). It is found that physical distances increase then locally decrease with increasing separation, with this transition occurring at ~2Mb, supporting the model of giant loops (Yokota *et al*, 2005).

The level of chromatin compaction varies across the genome, and can be observed by FISH. G-band, gene poor regions are more compact than R-band. Whilst both follow a random walk, the slope comparing mean square physical distance to genomic distance is less steep in G-bands than in R-bands. This is consistent with the higher compaction of chromatin from G-bands that can be observed with sucrose sedimentation, as the more compact G-bands move more quickly through a sucrose gradient than the less dense R-bands (Yokota *et al*, 1997; Gilbert *et al*, 2004). Indeed G-bands and R-bands form distinct compartments within the nucleus that are not able to interact with one another. Within those compartments, there are many more interactions between G-band chromatin, than between R-band chromatin (Lieberman-Aiden *et al*, 2009).

FISH assays can also detect changes in chromatin compaction at particular loci. It is found that chromatin compaction can vary as genes are differentially expressed during development (Chambeyron *et al*, 2004), and as a result of a mutation in a component of chromatin (Eskeland *et al*, 2010).

Another method of studying chromatin compaction is Chromosome conformation capture (3C), which detects the occurance of interactions between genomic regions (Figure 1.9). Cross-links between interactions are created using formaldehyde, then the genome is digested with a restriction enzyme. The formaldehyde links remain leaving pairs of short DNA fragments joined by cross-links. An intramolecular ligation step is carried out and the cross-links reversed. The ligation product can be detected by qPCR (quantitative polymerase chain reaction) using one primer from

each region of interest. This method is able to measure the interaction between two known regions (Dekker *et al*, 2002). To find the interaction between one known region and any number of unknown genomic regions 4C (circular 3C or 3C on chip) is used. Both ends of the fragment are ligated together giving a circular molecule. Using a pair of primers, one from each end of the region of interest, any regions that interact with the region of interest can be identified by sequencing, or by hybridising to microarrays (Zhao *et al*, 2006; Simonis *et al*, 2006). 5C (3C carbon copy) can also be used to study interactions on a large scale. A 3C library of cross-linked fragments is created and these are annealed to multiplex oligonucleotides. Oligonucleotides are designed to anneal to all predicted junctions, and can be either forward or reverse oligonucleotides. Forward oligonucleotides anneal to the 3' end of restriction fragments on the sense strand and have a T7 tag at the 5' end, and reverse oligonucleotides anneal to the 3' end of restriction fragments on the antisense strand and have a T3 tag at the 3' end. These then give a continuous oligonucleotide across the ligated ends which can then be ligated together using Taq ligase, forming a carbon copy library. The carbon copy library is then amplified using primers corresponding to the T7 and T3 tags, and can then be sequenced or studied by DNA microarrays (Dostie *et al*, 2006). 3C can be combined with ChIP analysis in 6C (combined 3C-ChIP-cloning), whereby 3C is carried out to the point of ligation, the ChIP is used to enrich for fragments containing the protein of interest. The fragments are then purified, reverse cross-linked and cloned into a vector. These can be screened by digestion, and fragments of interest are identified then verified by both 3C and ChIP (Tiwari *et al*, 2008).

More recently a technique called Hi-C has revealed the long-range interactions that occur across the genome. As with 3C, interacting DNA is crosslinked by formaldehyde, and cut up with restriction enzymes, however before ligation, ends are marked with biotin, which allows pull-down of the ligated fragments using Streptavidin beads. High throughput sequencing can then be used to identify interacting regions. Hi-C of the entire human genome confirmed the existence of CTs, the two distinct chromatin states of G and R bands and a hierarchical folding structure (Lieberman-Aiden *et al*, 2009; Figure 1.9).

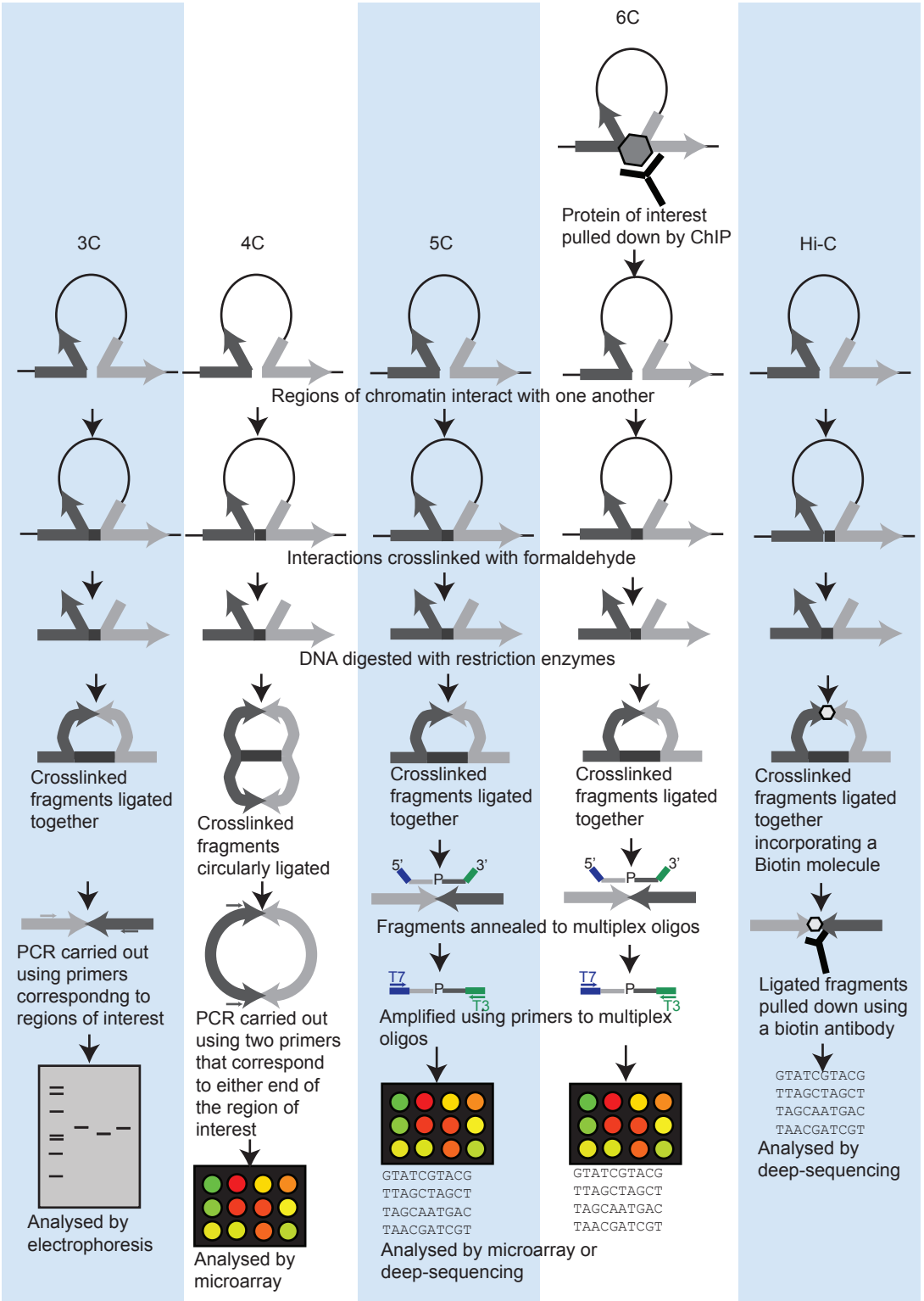


Figure 1.9 - 3C to Hi-C.
The C methods involve formaldehyde crosslinking of regions of the genome that interact with one another, cutting of the genome using restriction enzymes, ligation of those regions, then identification of those regions by various methods (Dekker *et al*, 2002; Zhao *et al*, 2006; Simonis *et al*, 2006; Dostie *et al*, 2006; Tiwari *et al*, 2008; Lieberman-Aiden *et al*, 2009).

1.5 Cohesin colocalises with CTCF on mammalian genome arrays

ChIP has been used to study cohesin localisation on mammalian genomes. Three independent studies found that cohesin colocalises with CCCTC-binding factor (CTCF) on chromosome arms (Figure 1.10; Wendt *et al*, 2008; Parelho *et al*, 2008; Stedman *et al*, 2008). It was found that 88% of CTCF sites are associated with cohesin and 79% of all cohesin binding sites can also associate with CTCF (Parelho *et al*, 2008). Moreover CTCF is shown to interact with SA1 (Rubio *et al*, 2008). The consensus binding sites of CTCF and cohesin are identical: CCACCAG(G/A)(G/T)GGC, the only difference being the probabilities of certain nucleotides (Parelho *et al*, 2008; Wendt *et al*, 2008) and this motif is essential for both CTCF and cohesin binding (Stedman *et al*, 2008). CTCF is essential for cohesin positioning, which is lost upon *CTCF* knockdown. There is conflicting evidence as to whether cohesin is required for positioning CTCF, as Wendt *et al* (2008) showed that knockdown of *Rad21* gave loss of CTCF positioning but Parelho *et al* (2008) demonstrated that it did not.

The interaction of cohesin and CTCF at the Kaposi's Sarcoma-associated Herpes-Virus (KSHV) has been more intensely studied through the cell cycle (Stedman *et al*, 2008; Kang and Lieberman, 2009). In lymphocytes CTCF and cohesin associate with cohesin/CTCF binding sites within the virus episome in its latent phase. From G_0/G_1 to M phase, both CTCF and the cohesin proteins accumulate, preferentially interacting with one another and with the KSHV binding sites in G_1-S, dissociating in G_2-M. An increase in KSHV latency associated transcripts from G_1 to G_2 is correlated with the accumulation of CTCF and cohesin proteins, and this cell cycle control is lost on mutation of CTCF binding sites in KSHV, suggesting that cohesin/CTCF binding is necessary for cell cycle regulation of gene expression. Association of cohesin with CTCF is able to repress the KSHV lytic genes, maintaining the latent state (Kang and Lieberman, 2009). It is unclear whether this cell-cycle dependant role of cohesin/CTCF in gene expression is linked to the role of cohesin in sister chromatid cohesion.

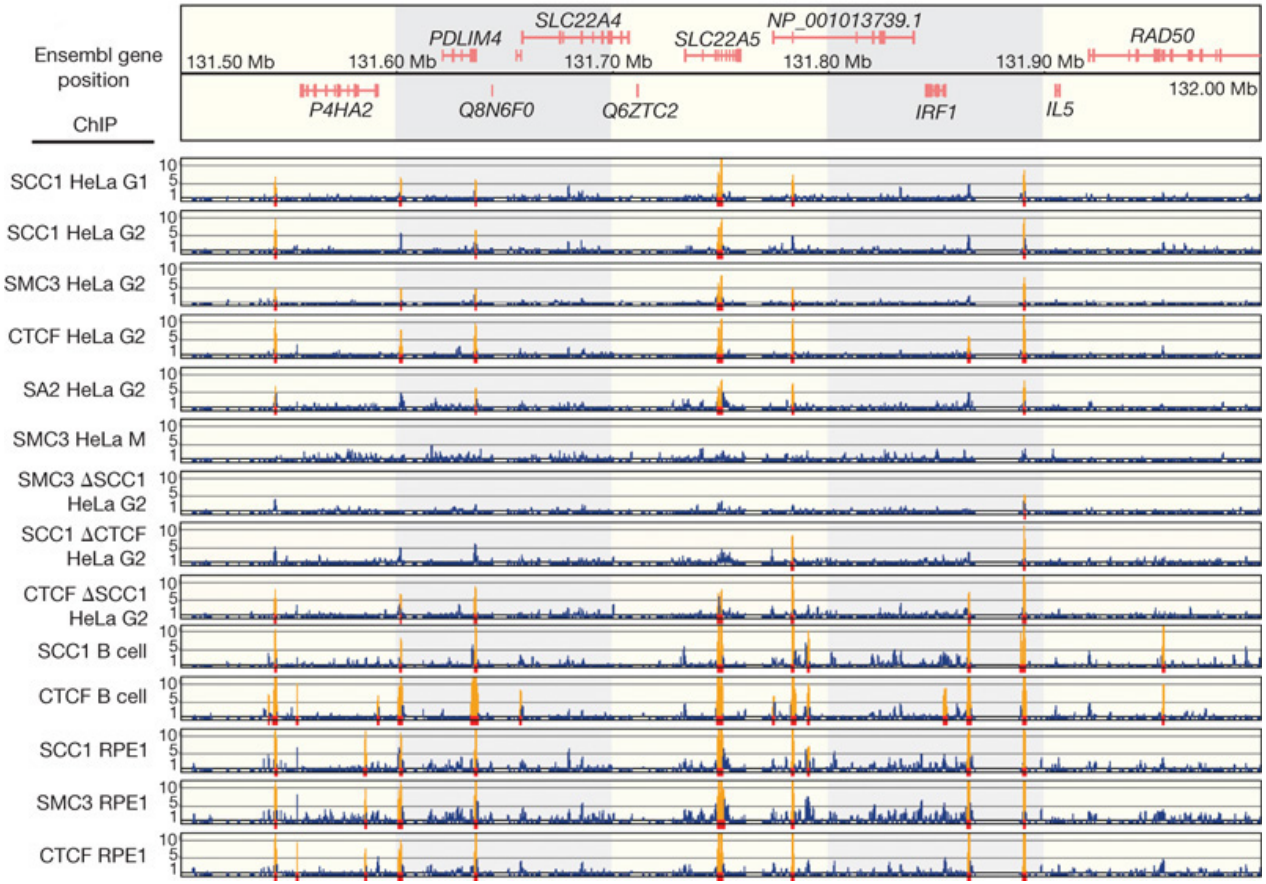


Figure 1.10 – Cohesin colocalises with CTCF.
ChIP data from Wendt *et al* (2008), showing that components of the cohesin complex colocalise with CTCF on chromosome arms in human cells. The data shown is from 131.5–132 Mb of human chromosome 5 on the Hg 18 assembly on Ensembl.

1.5.1 CTCF is an insulator protein

The previously discussed *gypsy* transposon is a well-known insulator in *Drosophila* (Parkhurst and Corces, 1986; Section 1.3.2). Like Suppressor of Hairy wing (Su(Hw)), the protein that interacts with the *gypsy* transposon in flies, CTCF is a Zinc-finger protein. CTCF is one of the best characterised insulator proteins in vertebrates.

The function of CTCF has been determined at the *H19/Igf2* imprinted region (Figure 1.11). *H19* and *Igf2* are expressed in a parent-of-origin specific manner due to imprinting. The locus consists of the *Igf2* gene, followed by the imprinting control region (ICR), which contains a CTCF binding site and two differentially methylated regions (DMRs), the *H19* gene and an enhancer. *Igf2* has a stronger promoter than *H19*, however both can only be transcribed with help from the long-range enhancer located downstream of *H19*. On the maternal allele CTCF binds to the ICR, blocking the enhancer from activating *Igf2*, allowing transcription of *H19*. On the paternal allele the ICR is methylated, preventing CTCF binding and allowing *Igf2* to interact with the enhancer. The strength of *Igf2*'s promoter is such that only *Igf2* is activated, and not *H19* (Kurukuti *et al*, 2006).

CTCF activity has been suggested to be dependant on a post-translational poly(ADP-ribosyl)ation. Insulator function is sensitive to inhibition of poly(ADP-ribose) polymerase by 3-aminobenzamide, in both CTCF reporter assays and at the *H19/Igf2* locus (Yu *et al*, 2004).

Much of the current data suggests that enhancers and insulators, such as CTCF, function through chromatin looping (Kurukuti *et al*, 2006; Splinter *et al*, 2006), however other possibilities have been proposed. For example it is suggested that CTCF may bind to particular nuclear sub-compartments via the nuclear matrix (Dunn *et al*, 2003). Co-purification of CTCF identified many proteins, including a strong interaction with nucleophosmin, a molecular chaperone involved in the transport of ribosome subunits and histones from the cytoplasm to the nucleus and nucleolus. ChIP showed that nucleophosmin binds to known CTCF insulated sites. A reporter transgene containing CTCF binding sites was studied using FISH and found to localise preferentially to the nucleolus. These findings suggest that CTCF may act

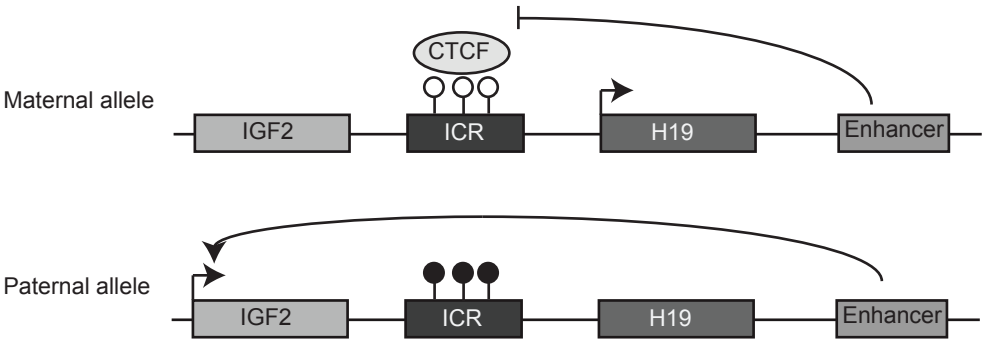


Figure 1.11– The imprinted *H19/Igf2* region.
On the paternal allele the ICR is methylated, blocking CTCF binding so that the enhancer can interact with the stronger promoter on *Igf2*, causing *Igf2* transcription. On the maternal allele, CTCF binds to the unmethylated ICR, and the enhancer is only able to interact with the closer *H19* promoter and only *H19* is transcribed. The lollipops indicate the DMRs within the ICR: filled circles indicate methyled CpGs and open circles indicate unmethylated CpGs.

by tethering DNA to the nucleolus (Yustufzai *et al*, 2004), however given how regularly CTCF binds along the genome, and that most of the genome is excluded from the nucleolus, this in unlikely to be the case *in vivo*. The epigenetic state of associated chromatin can also be altered by CTCF binding. CTCF associates with chromodomain helicase DNA-binding protein 8 (CHD8) at CTCF binding sites; RNAi knockdown of CHD8 results in loss of *Igf2/H19* imprinting and alteration of CpG methylation and histone acetylation adjacent to CTCF binding sites (Ishihara *et al*, 2006). The absence of CTCF binding can alter histone modifications at the *β-globin* locus, giving increased di-meH3K9 and di-meH3K27 (Splinter *et al*, 2006) and can lead to methylation of the ICR on the maternal allele of the *H19/Igf2* locus (Kurukuti *et al*, 2006).

A subpopulation of CTCF is suggested to associate with the largest subunit of RNA polymerase II (Rpol II), via the C-terminus of CTCF. They colocalise in the nucleus, through the genome, possibly associating at ACHs. The association of Rpol II with the ACH of the *H19/Igf2* locus was found to be dependant on CTCF binding site function in the ICR (Chernukhin *et al*, 2007). If this interaction does indeed take place, it can only be very transient, as ChIP data does not suggest colocalisation of CTCF and Rpol II on the genome (Barski *et al*, 2007).

By comparing CTCF/cohesin sites in B-cells and T-cells, it was seen that cohesin and CTCF associate with DNA in a cell type specific manner, which is dependant on DNA methylation, suggesting a mechanism by which CTCF/cohesin localisation might be maintained through cell division (Hadjur *et al*, 2009).

Genome wide CTCF binding sites have been mapped by ChIP (Barski *et al*, 2007; Kim *et al*, 2007; Bao *et al*, 2007). Taking every "hit" from ChIP experiments, CTCF binds to the genome at an average density of 906 sites/Mb. However when a binding threshold is applied to this data, the average density is 4.6 sites/Mb (Wendt *et al*, 2008).

It is likely that cohesin and CTCF act together to alter aspects of higher order chromatin structure to mediate insulation and enhancement in non-replicating cells. CTCF may create interactions between binding sites and cohesin may hold them in place. For example the ACH seen at the *H19/Igf2* and *β-globin* loci may begin as

interactions between CTCF and its binding sites, but they are secured by cohesin rings (Nativio *et al*, 2009).

1.5.2 Cohesin and CTCF contribute to DNA looping

The effect of cohesin and CTCT on chromatin looping has been studied at a number of locations in the human genome using 3C. In the older data, CTCF is shown to be necessary for DNA looping and correct gene expression (Zhao *et* al, 2006; Kurukuti *et al*, 2006; Splinter *et al*, 2006), whereas more recent studies include data identifying cohesin as equally necessary for looping and gene expression (Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010).

CTCF has been implicated in looping at the *β-globin* locus, which contains a number of developmentally regulated erythroid specific genes under control of an upstream locus control region (LCR). Three CTCF binding sites have been identified upstream of the locus (HS-85, HS-62 and HS-5) and one downstream (3'HS-1) in mice. 3C data shows that in erythroid cells, the four CTCF binding sites spatially interact with the LCR and the active *β-globin* genes (Splinter *et al*, 2006). Further data suggests that long-range interactions at this region are cell type specific, with a more densely clustered organisation in the absence of *β-globin* activity. These interactions are dependant on both cohesin and CTCF. Knockdown of CTCF in erythroid cells, which express globin genes, leads to acquisition of repressive histone marks at the *β-globin* locus and loss of *β-globin* expression (Hou *et al*, 2010). It seems counterintuitive that loss of CTCF should lead to loss of gene expression, when non-expressing cells have denser CTCF interactions, but it is possible that repressive histone marks act as a fail-safe, to knockdown gene expression in the absence of the repressive CTCF loops.

More recent data is available for the imprinted *H19/Igf2* region (Figure 1.11). RNAi experiments show that both CTCF and cohesin are essential for imprinting at the *H19/Igf2* locus (Figure 1.11; Wendt *et al*, 2008; Parelho *et al*, 2008; Stedman *et al*, 2008; Nativio *et al*, 2009). There are 114 intra- and interchromosomal interactions with the maternal *H19/Igf2* ICR (Zhao *et al*, 2006). A 3C study found

that at the imprinted *H19/Igf2* region, there was no enhancer access to the maternal *Igf2* allele, whilst the paternal allele was accessible, however there was no difference in enhancer access to the *H19* promoter between the methylated paternal allele and the unmethylated maternal allele. The ICR remains in close physical proximity to all regions in the *Igf2/H19* domain on both the maternal and paternal allele, except the paternal DMR1, maternal P1/DMR2 and paternal MAR3, which are excluded from the ICR. This gives evidence of higher order looping that may regulate gene expression at this locus, bringing the active genes into an ACH containing regulatory elements and genes (Kurukuti *et al*, 2006). Another 3C study confirmed the existence of these interactions and found that looping at the *H19/Igf2* region was abrogated upon Rad21 knockdown by RNAi. This effect was tested in cells synchronised in G_1 and in G_2 by double thymidine block, demonstrating that cohesin mediates looping independently of the cell cycle (Nativio *et al*, 2009). It is interesting to note that, even though these studies indicate that cohesin appears to be essential for imprinting, no imprinting phenotypes, such as those seen in Beckwith-Wiedemann syndrome, are observed in CdLS. This questions the biological role of cohesin, or at least NIPBL, in imprinting *in vivo*.

At the Interferon γ (IFNγ) locus, CTCF and cohesin are also shown to mediate loops that affect gene expression. In Th2 cells, which do not express IFNγ, far less cohesin and CTCF binds at the three cohesin/CTCF binding sites, upstream, intragenic and downstream of the IFNγ locus, compared to Th1 cells, which do express IFNγ. A 3C assay shows that in Th1 cells, but not in Th2, these three binding sites interact with one another, and these interactions are abrogated on RNAi knockdown of CTCF or Rad21. Knockdown of CTCF results in loss of cohesin localisation at these sites, whilst CTCF localisation was retained after Rad21 knockdown, however there was decreased IFNγ expression. These data suggest that in Th1 cells, CTCF localised cohesin to these sites, forming loops which are required for optimal expression of IFNγ (Hadjur *et al*, 2009).

Similarly, CTCF and cohesin binding sites were identified at the apolipoprotein (APO) gene cluster, which contains genes essential for the metabolism and redistribution of lipids and lipoproteins. Like the IFNγ locus, the APO cluster has three CTCF binding sites, which also bind cohesin. 3C found that these sites are able

to interact with one another, and that these interactions are dependant on CTCF and cohesin. Knockdown of CTCF or cohesin also reduced the transcription of *APOC3* and increased the transcription of *APOA1* and the surrounding genes. It is thought that CTCF and cohesin hold the APO cluster in a loop conformation that facilitates transcription of *APOC3*, but suppresses transcription of *APOA1* (Mishiro *et al*, 2009).

Cohesin and CTCF have been shown to have a role in genome rearrangement at the *Igh* locus, which encodes the heavy chain of antibodies, and the *Igκ* and *Igλ* loci, which encode the two forms of the light chain of antibodies. All of these loci undergo genomic recombination, giving the many variants of antibodies that are necessary for the acquired immune response. At *Igκ* and *Igλ*, a significant contraction occurs at the appropriate time for genome rearrangement in pro-B cells, and this correlates to stage specific recruitment of cohesin and CTCF to the loci. While this is a different effect to the gene expression role at other loci, there is no reason to consider that the same mechanism, of looping mediated by CTCF and cohesin, is not responsible for this contraction (Degner *et al*, 2009). However, no immune deficiencies have been observed in CdLS patients.

It is reasonable to consider that the specific interactions that have been studied at the *β-globin* locus, the imprinted *H19/Igf2* region, the IFNγ locus and the APO gene cluster and the *Igh*, *Igκ* and *Igλ* loci are representative of other regions of the genome. Throughout the genome interactions such as this may be speculated to take place, and in CdLS even a minor disruption of these interactions may give the general low-level dysregulation of gene expression that is observed in this disease (Liu *et al*, 2009).

1.5.3 CTCF and cohesin can regulate gene expression by insulation

CTCF and cohesin have also been found to colocalise at intragenic chromatin boundaries (Gomes and Espinosa, 2010). The PUMA (p53 upregulated modulator of apoptosis) locus, which orchestrates programmed cell death upon upregulation by p53, is found to constituently produce a 6kb unprocessed non-coding transcript in the

absence of p53, which presumably represses PUMA expression. The 6kb locus that this RNA is transcribed from is marked by active histone marks including H3K4me3 and H3K9Ac, and is flanked by repressive H3K9me3 marks. CTCF and cohesin binding is also found across the 6kb locus and in the absence of CTCF, PUMA expression is increased and the chromatin marks on this region are lost. It is considered that CTCF and cohesin somehow maintain the chromatin state of this region, giving the continual expression of the 6kb transcript, required to repress PUMA.

1.6 Cohesin and NIPBL associate with transcription factors without CTCF

Evidence also suggests that cohesin and NIPBL can affect gene expression independently of CTCF. Cohesin and NIPBL have been shown to associate with transcription factors and components of the transcription machinery in a variety of tissues to regulate gene expression via chromatin looping (Kernohan *et al*, 2010; Schmidt *et al*, 2010; Kagey *et al*, 2010).

At the *H19*/*Igf2* and *Gtl2*/*Dlk1* loci, ATRX (α-Thallassaemia mental retardation X-linked) and MeCP2 (methyl CpG binding protein 2) are shown to be necessary for silencing of imprinted genes (Kernohan *et al*, 2010; Cunningham *et al*, 2010). In post-mitotic cells in the mouse brain, ATRX and MeCP2 are found to co-immunoprecipitate with SMC1 and SMC3, and both are found to colocalise with cohesin at the *H19*/*Igf2* locus. At this locus, ATRX and MeCP2 associate not with the methylated paternal ICR, but with the unmethylated CTCF/cohesin-bound maternal ICR. At the *Gtl2*/*Dlk1* locus, ATRX and MeCP2 localised adjacent to CTCF/cohesin sites. Like *H19*, *Gtl2* is maternally expressed as CTCF binds to the DMR, and *Dlk1*, like *Igf2*, is expressed from the paternal allele where the DMR is methylated. The interaction of these proteins in silencing imprinted genes may explain the small overlap in mental retardation phenotypes between α-Thallassaemia mental retardation X-linked (caused by mutation of ATRX), Rett syndrome (caused by mutation of MeCP2) and CdLS.

A ChIP-seq study in breast and liver cancer cells found that where cohesin binds in the absence of CTCF, it binds instead with a tissue specific global regulator (Schmidt *et al*, 2010). Many oestrogen regulated genes in MCF-7 breast cancer cells bound to both cohesin and oestrogen receptor alpha (ER), but not CTCF, and these regions are enriched for interchromosomal loop anchors. Knockdown of cohesin prevented re-entry into the cell cycle when cells were stimulated with oestrogen, however it is not clear if this is due to loss of sister chromatid cohesion or a problem with oestrogen stimulated gene expression. Although this gene regulation by cohesin is independent of CTCF, it appears to occur by a similar mechanism, by tethering loops of DNA together that may have been created by another protein, in this case ER. Similarly, in liver cancer cells, cohesin was found to associate with the liver-specific transcription factors, HNF4A and CEBPA, at different genomic locations.

Most recently, genome wide studies in mouse ES cells have found that cohesin and Nipbl associate with the Mediator complex (Figure 1.12; Kagey *et al*, 2010). Mediator is a global regulator of transcription that interacts with RpolII. The complex is 1.2MDa and has 26 subunits, which change in structure and composition depending upon the context of the promoter. It is this versatility that is thought to allow Mediator to regulate gene expression in such a global and yet tissue-specific manner (Taatjes, 2010). Mediator, Nipbl and cohesin were shown to co-precipitate and co-localise at enhancers and promoters, promoting the formation of loops between enhancers and promoters, as shown by 3C. These interactions were specific to ES cells, and associated with ES-cell specific genes. Mediator did not associate with all cohesin binding sites, and its association was mutually exclusive with CTCF binding. Interestingly Nipbl only associated with Mediator-cohesin binding sites, and not with CTCF-cohesin binding sites suggesting that Mediator-cohesin mediated looping is a far more significant contributor to CdLS than CTCF-cohesin mediated looping. Mediator and cohesin co-occupied different sets of promoters in ES cells and MEFs, unlike cohesin and CTCF, which co-occupied the same sites in both cell lines, suggesting that Mediator and cohesin regulate expression of tissue specific genes, whereas CTCF and cohesin regulate more housekeeping genes.

These data, taken together, suggest that CTCF may colocalise with cohesin on chromatin to regulate expression of some genes. However Mediator and cohesin,

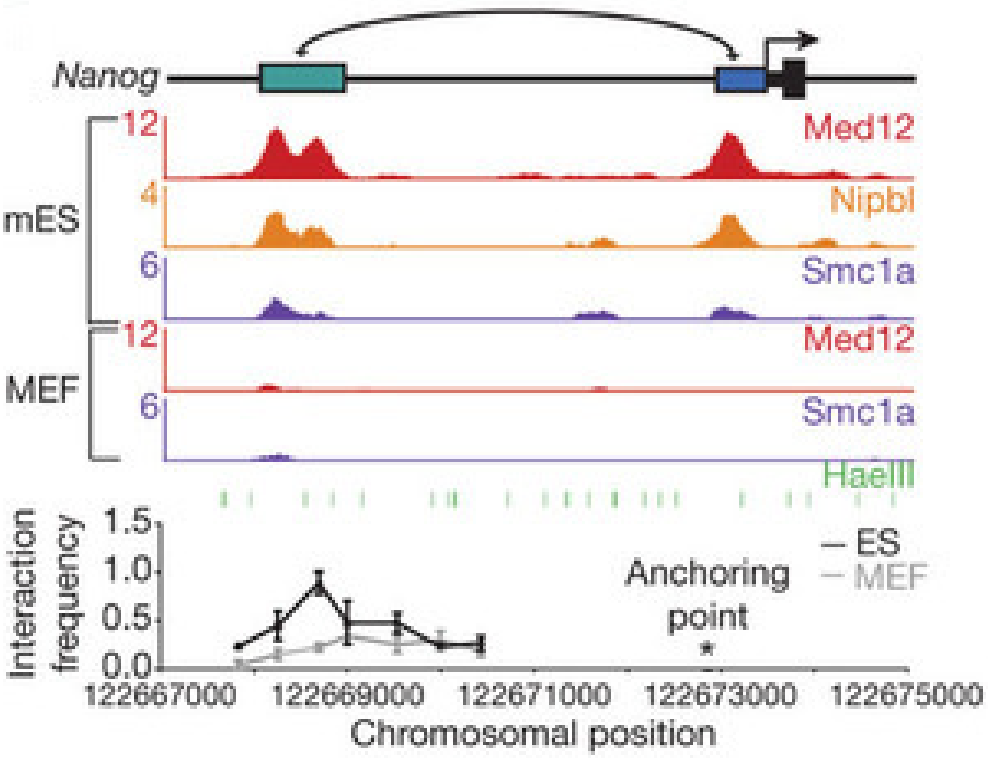


Figure 1.12 - Mediator, cohesin and Nipbl colocalise on mouse chromatin to regulate chromosome topology.
Data showing the interaction of Mediator, coheisn and Nipbl at the Nanog locus in mouse ES cells (from Kagey *et al*, 2010). A schematic showing the Nanog locus is shown at the top with ChIP-seq data showing the binding of Med12, Smc1a and Nipbl in ES cells and MEFs (shown in reads per million with the y axis floor set to 0.5 reads per million). The site of restrictions enzymes is shown below, with a 3C graph at the bottom (error bars represent the standard error of the average of 3 independent PCR reactions). The genomic coordinates are build NCBI36/mm8.

interacting with tissue specific transcription factors, may regulate certain tissue specific genes.

1.7 The effect of cohesinopathies on chromatin conformation

1.7.1 Yeast cohesinopathy models have defects in chromatin conformation

Models of the cohesinopathies have been made in budding yeast, introducing CdLS and RBS associated mutations into the endogenous cohesin proteins, Scc2, Smc1 and Eco1. Nuclear organisation was studied in these mutants (Gard *et al*, 2009). The morphology of the nucleolus of the CdLS and RBS mutants was altered. In yeast, the telomeres and centromeres are tethered to the nuclear periphery, and some genes are able to relocate within the nucleus upon activation. In the CdLS mutants neither the centromere nor telomere localisation was affected, however in the RBS mutants telomeres clustering was lost.

The *GAL2* gene is normally localised close to the nucleolus, as it is genomically close to the rDNA, but moves to the nuclear periphery upon activation, still remaining close to the nucleolus. *GAL2* also contains a cohesin binding site that alters upon activation. In the cohesinopathy mutants, *GAL2* loses its colocalisation with the nucleolus, and when induced it is massively overexpressed. General chromatin compaction was also measured in these yeast models by introducing pairs of LacO operators into the genome, a known distance apart. These could then be visualised in live cells by introducing LacI repressor fused with GFP, and the distance between pairs determined as a measure of chromatin compaction. It was found that in CdLS and RBS models chromatin tended to be less condensed than wildtype.

The tRNA genes (transcribed by Rpol III) cluster near the nucleolus in yeast, and can silence neighbouring Rpol II transcribed genes, known as tRNA gene-mediated (tgm) silencing. The CdLS and RBS mutations alleviated tgm-silencing of an adjacent gene, and FISH demonstrated that in the cohesinopathy mutants, the tRNA

genes did not cluster by the nucleolus. It is known that condensin mutants also disrupt tRNA gene localisation and alleviate tgm-silencing, and both CdLS and RBS mutations disrupted condensin loading, so it is unclear whether this is a direct or indirect effect. However while it is unsurprising that mutation of Scc2, which also loads condensin, affected condensin binding, it is interesting that Eco1 mutation also affected condensin.

1.7.2 Possible mechanisms for altered chromatin conformation in cohesinopathies

A yeast two-hybrid experiment showed that NIPBL interacts with HDAC1 and HDAC3, and that NIPBL is able to recruit the HDACs to a reporter construct. This effect was abrogated with CdLS mutations of NIPBL, and with treatment by the histone deacetylase inhibitor TSA. Histone acetylation was also shown to be increased in NIPBL mutants (Jahnke *et al*, 2008). It is known that histone modifications are able to affect the condensation of chromatin, and acetylation of H4-K16 results in the decompaction of chromatin (Tse *et al*, 1998; Carruthers and Hansen, 2000; Tóth *et al*, 2004; Shogren-Knaak *et al*, 2006; Robinson *et al*, 2008). However, as the link has, so far, only been shown between NIPBL and HDACs, this does not explain any changes in chromatin condensation in SMC1L1 or ESCO2 mutants, as observed in yeast RBS and CdLS models (Gard *et al*, 2009).

NIPBL does not only load cohesin, but, in *S. cerevisiae*, Scc2 also loads condensin, which is involved in compacting the chromatin prior to metaphase (D'Ambrosio *et al*, 2008). One could hypothesise that condensin may also compact chromatin throughout the cell cycle, however there is not yet any evidence to suggest this.

There is plenty of evidence to suggest that cohesin and CTCF may work together to mediate intra-chromosomal interactions (Wendt *et al*, 2008; Hadjur *et al*, 2009; Parelho *et al*, 2008; Stedman *et al*, 2008; Degner *et al*, 2009; Nativio *et al*, 2009; Mishiro *et al*, 2009), and that cohesin may work with other transcription factors including Mediator in the same way (Kagey *et al*, 2010; Schmidt *et al*, 2010). Loss of these tightly mediated loops may result in a general disorganisation of chromatin,

which could also explain the effect of CdLS and RBS mutations on chromatin in yeast models.

1.8 Aims

In this thesis I have set out to explore the behaviour of chromatin in CdLS by analysing its compaction using the FISH assay. I have begun to analyse the role of the different proteins involved in this compaction, including CTCF, NIPBL and cohesin itself using RNAi. I have also considered the differing effects between different species and cell lines.

2 Materials and Methods

2.1 Reagents, stock solutions and buffers

Acetate buffer: 5M potassium acetate and 11.5% glacial acetic acid in dH_2O.

Antibiotics: Stock solutions of antibiotics were made up at 1000 X. Chloramphenicol was made up to 25mg/ml in ethanol (EtOH).

Citrate buffer: A 1l stock was made up by dissolving 85.5g sucrose and 11.76g of trisodium citrate in 800ml dH_2O with 50ml DMSO, adjusted to pH 7.6 with 2-3 drops of concentrated HCl. This was made up to 1l with dH_2O and could be stored at 4°C for up to 2 years.

Cold lysis buffer: 0.2N NaOH with 1% SDS in dH_2O, made fresh for each use.

Diethylpyrocarbonate (DEPC) treated water: dH_2O is treated with 0.1% (w/v) DEPC for 1 hour at 37°C then autoclaved.

DNase I Buffer: Supplied as a 10 X buffer, made up of 100mM Tris-HCl pH 7.5, 25mM $MgCl_2$ and 5mM $CaCl_2$, from Ambion.

Elutriation buffer: PBS with 1% FCS, 0.3mM EDTA and 0.1% glucose.

FACS Solution A: 20µg/ml Trypsin Type IX-S (T0303, Sigma) in Stock Solution pH 7.6, stored at -20°C.

FACS Solution B: 0.5mg/ml Trypsin Inhibitor (T9253, Sigma) and 0.1mg/ml RNAse A (R4875, Sigma) in Stock Solution pH 7.6, stored at -20°C.

FACS Solution C: 416µg/ml Propidium Iodide (81845, Sigma) and 1mg/ml spermine tetrahydrochloride (S2876, Sigma) in Stock Solution pH 7.6, stored at -20°C.

FACS Stock Solution: Dissolve 1g/l Trisodium Citrate, 60.5mg/l Tris Base, 522mg/l Spermine Tetrahydrochloride and 0.1% Igepal (v/v) in dH_2O adjusted to pH 7.6.

GTE buffer: 50mM glucose, 25mM Tris-HCl pH 8.0 and 10mM EDTA pH 8.0 in dH_2O. Just prior to use, 2ml of buffer was aliquoted and a pinch of lyzozyme (Sigma) added.

Hybridisation mix: 50% deionised formamide, 10% dextran sulphate and 1% Tween 20 in 2 X SSC, made fresh each time.

Luria-Bertani (LB) Agar: 10g of tryptone, 5g of yeast extract, 10g of NaCl and 15g agar was added to 1 litre of dH_2O. NaOH was added to bring the solution to pH 7.0. The solution was then autoclaved. Prepared by technical services at the MRC Human Genetics Unit.

Luria-Bertani (LB) Broth: 10g of tryptone, 5g of yeast extract and 10g of NaCl was added to 1 litre of dH_2O. NaOH was added to bring the solution to pH 7.0. The solution was then autoclaved. Prepared by technical services at the MRC Human Genetics Unit.

MAA: Methanol and acetic acid in a ratio of 3:1, prepared fresh each time.

Nick Translation Salts: 0.5M Tris pH 7.5, 0.1M $MgSO_4$, 1mM DTT and 0.5mg/ml BSA fraction V (Sigma), prepared by Shelagh Boyle.

Orange G loading buffer: 5 X loading buffer was made up of 10% (w/v) sucrose and 0.1% (w/v) Orange G in dH_2O.

PCR buffer: 200mM Tris-HCl (pH 8.4) and 500mM KCl in dH_2O, prepared by Invitrogen.

Phenol-chloroform: Phenol:chloroform:isoamyl alcohol 25:24:1 was obtained from Sigma.

Phosphate Buffered Saline (PBS): Made up of 10mM phosphate, 137mM NaCl and 27mM KCl. Prepared from tablets purchased from Unipath by technical services at the MRC Human Genetics Unit.

SDS loading buffer: Was prepared as a 4 X stock solution of 250mM tris HCL (pH 6.5), 8% sodium dodecyl sulphate (w/v), 20% β-mercaptoethanol (v/v), 40% glycerol (v/v) and 0.1% bromophenol blue (w/v).

SDS-PAGE running buffer: 25mM Tris base, 250mM glycine (pH 8.3) and 0.1% sodium dodecyl sulphate (w/v) was prepared as a 5 X stock.

SiRNA buffer: 60mM KCl, 6mM HEPES pH 7.5 and 0.2mM $MgCl_2$ treated with 0.1% (w/v) DEPC for 1 hour at 37°C then autoclaved.

SSC: 3M NaCl and 0.3M tri-sodium citrate at pH 7.4. Prepared as a 20 X stock by technical services at the MRC Human Genetics Unit.

SuperScript™ II First Strand Buffer: 5 X buffer 250mM Tris-HCl, 375mM KCl and 15mM $MgCl_2$. Supplied by Invitrogen.

TE: 10mM Tris HCL (pH 7.6) and 0.1mM EDTA in dH_2O and autoclaved. Prepared by technical services at the MRC Human Genetics Unit.

Tris Acetate Buffer, TAE: A 50 X stock solution was made up of 2M Tris acetate and 50mM EDTA (pH 8.0), prepared by dissolving 242g Tris Base in 750ml dH_2O, adding 57.1ml glacial acid acid and 100ml of 0.5M EDTA (pH 8.0) and making up to 1l with dH_2O. Prepared by technical services at the MRC Human Genetics Unit.

Tris Borate Buffer, TBE: A 20 X stock solution was made up of 90mM Tris Borate, 2mM EDTA (pH 8.0), prepared by dissolving 108g of Tris Base and 27.5g of Boric Acid in 400ml of 0.5M EDTA and 960ml dH_2O. Prepared by technical services at the MRC Human Genetics Unit.

Tris Buffered Saline (TBS): 150mM NaCl and 10mM Tris-HCl (pH7.4) in dH_2O.

TRIZOL® Reagent: TRIZOL® reagent contains phenol and guanidine isothiocyanate and is supplied by Invitrogen.

Western Transfer Buffer: 25mM Tris-Glycine (pH 8.3) with 20% methanol (v/v) was made fresh before each use by dissolving 3.03g of Tris base and 14.4g of glycine in 800ml of dH_2O, then 200ml of methanol were added.

2.2 Plasmids and bacterial culture

2.2.1 Genomic clones

Genomic clones were supplied by either the Sanger Institute Clone service (service no longer available; http://www.sanger.ac.uk/), or by the BacPac Resources Centre at the Children's Hospital Oakland Research Institute (http://bacpac.chori.org/). Fosmid clones were used for all experiments, which came from the WIBR2 library derived from a human female at the Broad Institute cloned into a pEpiFOS-5 vector, in DH10B T1 resistant competent cells. Clones were supplied as bacterial stab cultures in agar. The fosmid clones used are listed in Table 2.1.

2.2.2 Bacterial culture

Bacteria were streaked out onto LB-agar plates containing a selective antibiotic. The LB-agar was melted in a microwave and the antibiotic added after it had cooled slightly, then poured into Petri dishes, ~20ml per 10cm dish. Bacteria from stab cultures or frozen glycerol stocks were streaked out to give single colonies and these were left to grow overnight at 37°C. For growth of fosmid clones, plates were supplemented with 25μg/ml chloramphenicol.

To prepare small quantities of DNA from bacteria, a single colony was used to inoculate 5ml of LB-broth, using the same selective antibiotic(s) as used in the LB-agar. They were then allowed to grow overnight with shaking at 37°C with at least a 5:1 air to liquid ratio.

2.2.3 Bacterial glycerol stocks

Glycerol stocks were prepared by adding 30% (v/v) glycerol to a 1ml aliquot of an overnight culture. Glycerol stocks are frozen at -70°C.

Table 2.1 - Fosmid probes used for FISH

Probe positions are taken from the GRCh37/h19 assembly on the UCSC genome browser (Kent et al, 2002; http://genome.ucsc.edu/)

Whitehead name	Other name	Region	Source	Start position	End position	Size (bp)	Role
WI2-671I21	G248P80020E11	11q13	Sanger	64768160	64811897	43737	Reference probe
WI2-1737E8	G248P86034C4	11q13	Sanger	65019283	65058937	39654	250kb probe
WI2-412A22	G248P81370A11	11q13	Sanger	65285668	65329285	43617	500kb probe
WI2-2372A13	G248P82828A7	11q13	Sanger	65519199	65559973	40774	750kb probe
WI2-3325E18	G248P800138C9	11q13	Sanger	65764738	65807012	42274	1Mb probe
WI2-2080I19	G248P87792E10	EDC	Sanger	151830653	151872769	42116	Reference probe
WI2-2423D6	G248P80671B3	EDC	Sanger	151569368	151609225	39857	250kb probe
WI2-553L9	G248P80305F5	EDC	Sanger	151280817	151318413	37596	500kb probe
WI2-3057F14	G248P800827C7	EDC	Sanger	151091201	151129333	38132	750kb probe
WI2-784N16	G248P81347G8	EDC	Sanger	150776620	150816117	39497	1Mb probe
WI2-3163O19	G248P8056H10	HoxD 2q31	Sanger	176638573	176678511	39938	Reference probe
WI2-1038D9	G248P82453B5	HoxD 2q31	Sanger	176890476	176932394	41918	250kb probe
WI2-2744M3	G248P8196G2	HoxD 2q31	BacPac	177034650	177074781	40131	375kb probe
WI2-1919C15	G248P87408B8	HoxD 2q31	Sanger	177130268	177168594	38326	500kb probe
WI2-1957H23	G248P85797D12	HoxD 2q31	Sanger	177381571	177420055	38484	750kb probe
WI2-1241H8	G248P82115D4	HoxD 2q31	Sanger	177647644	177682721	35077	1Mb probe
WI2-1702P7	G248P87869H4	18q22	BacPac	64134988	64175629	40641	Reference probe
WI2-502C21	G248P8988B11	18q22	BacPac	63893119	63939831	46712	250kb probe
WI2-1735D16	G248P85999B8	18q22	BacPac	63759712	63799511	39799	375kb probe
WI2-1708L7	G248P87905F4	18q22	BacPac	63648826	63690385	41559	500kb probe
WI2-920O6	G248P80414H3	Protocadherin chr5	BacPac	140125390	140167323	41933	Reference probe
WI2-812D24	G248P80979B12	Protocadherin chr5	BacPac	140888956	140927951	38995	750kb probe
RP1-61M11		11p13	ICRF RLDB	32146946	32226183	79237	Non-looping probe

2.2.4 Preparation of DNA from bacterial overnight cultures

Fosmid DNA was prepared by alkaline lysis mini-prep. ~1.5ml cultures were pelleted by centrifugation at 16,000*g* for 30s, and were resuspended in 200µl GTE buffer for 5min. 400µl cold lysis buffer was added and left on ice for 5min before 300µl acetate buffer was added to precipitate the cell debris. After 5min on ice the flocculent precipitate was spun down at 4°C for 5min at 16,000*g*, and the supernatant combined with an equal volume (~800µl) of phenol-chloroform to clean. This was spun down at 4°C for 2min at 16,000*g* and the aqueous top layer removed and mixed with an equal volume of isopropanol (~800µl), and kept for one hour at ~20°C. The DNA was pelleted by 15min centrifugation at 16,000*g* at 4°C, then washed in 70% EtOH, spun down again for 5min at 4°C at 16,000*g* then resuspended in 30µl TE.

2.3 Preparation and handling of DNA

2.3.1 Quantification of DNA by spectrophotometry

DNA concentration was checked at every stage of experiments by spectrophotometry, on a nanodrop ND-1000 (Nanodrop technologies) according to the manufacturer's instructions. The absorbance of the sample was measured at a wavelength of 260nm (A_{260}); 50µg/ml of double stranded DNA will give an A_{260} measurement of 1.

2.3.2 Resolution of DNA on agarose gels

For routine DNA analysis, gels were made with an appropriate agarose percentage ("Hi-Pure" Low Eeo agarose, BioGene UK, w/v) in Tris-Borate Buffer (TBE), with 0.5µg/ml Ethidium Bromide (EtBr). DNA was made up with Orange G loading buffer, with 1 part buffer added to 4 parts sample. Appropriate commercial DNA size markers were used to allow size determination and quantification of size fragments. After resolution, DNA was visualised under UV illumination. For analysis gels were scanned using a 532nm laser and a 575nm Long Pass filter (LPG filter on a Fuji FLA-5100 phosphoimager).

2.4 Preparation and handling of RNA

2.4.1 Quantification of RNA by spectrophotometry

RNA concentration was determined using spectrophotometry, on a nanodrop ND-1000 (Nanodrop technologies) according to the manufacturers instructions. The absorbance of the sample was measured at a wavelength of 260nm (A_{260}); 40µg/ml of RNA will give an A_{260} measurement of 1.

2.4.2 RNA isolation and purification

RNA was isolated from cells using TRIZOL® reagent (Invitrogen), according to the manufacturer's instructions. Cells in suspension were pelleted by centrifugation at 1200*g* for 4min. 1ml per 5-10 X 10^6 cells of TRIZOL® Reagent was added and mixed by pipetting to lyse the cells. The cells were homogenised for 5min at room temperature, then 200µl chloroform was added to separate the mixture into phases. The tubes were mixed well and incubated at room temperature for 2-3min, before centrifugation at 12,000*g* for 15min at 4°C. This separated the mixture into three phases, a red phenol-chloroform phase, a small white phase and a clear aqueous phase that contained the RNA. The RNA containing phase was transferred to a clean tube, and isopropanol added, around half the original volume of TRIZOL® Reagent. RNA was precipitated at room temperature for 10min and pelleted by 10min centrifugation at 4°C at 12,000*g*. The pellet was washed in 75% EtOH made with Diethylpyrocarbonate (DEPC) treated dH_2O, the same volume as the volume of TRIZOL® Reagent used, and centrifuged at 7,500*g* for 5min at 4°C. The pellet was then dissolved in 25µl DEPC treated dH_2O and the RNA concentration estimated using a Nanodrop.

2.4.3 DNase treatment of RNA

RNA was DNase treated using the Ambion DNA-*free*™ kit, according to the manufacturer's instructions. 3µl 10 X DNase I buffer was added to the RNA

solution along with 1µl rDNase I, and mixed gently. The reaction was incubated at 37°C for 30min, then stopped by adding 3µl DNase Inactivation Reagent and incubating for 2min at room temperature. This was then centrifuged at 10,000*g* for 1min 30s and the supernatant transferred to a fresh tube. The RNA concentration was then estimated using a Nanodrop.

2.4.4 Resolution of RNA on agarose gels

For routine RNA analysis, gels were made with an appropriate agarose percentage ("Hi-Pure" Low Eeo agarose, BioGene UK, w/v) in Tris-Acetate Buffer (TAE), with 0.5µg/ml ErBr. RNA was made up with Orange G loading buffer and 50% formamide, then denatured at 65°C for 10min and cooled on ice for 5min. Appropriate commercial RNA size markers were used to allow size determination and quantification of size fragments. After resolution, RNA was visualised under UV illumination. For analysis, gels were scanned using a 532nm laser and a 575nm Long Pass filter (LPG filter on a Fuji FLA-5100 phosphoimager).

2.4.5 Reverse transcription polymerase chain reaction (RT-PCR)

cDNA was prepared using a modified Invitrogen SuperScript™ II protocol. For each reaction, 4µl of Random Primers (Promega) were mixed with 1µg RNA and made up to 11.5µl with dH_2O. This was denatured for 5min at 65°C, then cooled for 5min on ice. 1µl 10mM dNTP mix, 4µl First-Strand buffer (Invitrogen), 2µl 0.1M DTT (Invitrogen) and 0.5µl RNase inhibitor (Invitrogen) were added to each tube, and incubated for 2min at room temperature. For each sample, a reverse transcriptase positive (RT+) test and a reverse transcriptase negative (RT-) control was set up; 1µl SuperScript™ II reverse transcriptase was added to the RT+ tubes, and 1µl dH_2O added to the RT- tubes. These were incubated at room temperature for 10min, at 42°C for one hour and inactivated at 70°C for 15min.

PCR was carried out using 1 X PCR buffer (Invitrogen), 2.5µM $MgCl_2$, 0.5µM of each primer, 0.2mM of each dNTP and 0.2 units Platinum Taq (Invitrogen), with 1µl

cDNA to give a total volume of 25µl. PCR reactions were carried out on a Peltier Thermal Cycler 225 from MJ Research. The PCR protocol used was:

1. 95°C for 10min
2. 95°C for 30s
3. 58°C for 30s
4. 72°C for 30s
5. Return to step two, *number* (Table 2.2) times
6. 72°C for 3min

Quantitative PCR (qPCR) was carried out on a Roche Lightcycler 480, using Quantitect® SYBR® Green from Qiagen. 1µl cDNA was made up with 1 X Quantitect® SYBR® Green PCR master mix and 0.3µM of each primer to give a total volume of 50µl. The qPCR protocol was:

1. 95°C for 15min
2. 95°C for 15s
3. 56°C for 30s
4. 72°C for 30s with a single acquisition of data
5. Return to step two, 49 times
6. 95°C for 5s
7. 65°C for 1min
8. Heat to 97°C at a rate of 0.29°C/s, acquiring data every 0.5°C
9. 40°C for 30s

The relative gene expression was calculated by comparison to expression of GAPDH.

The primers used for RT-PCR were purchased from Sigma, and are detailed in Table 2.2.

Table 2.2 - Primers used for RT-PCR

Primer name	Sequence	Melting temperature (^{o}C)	Gene	Product Length	PCR cycles used
Nip1F	5'-CTGCGGAGTGACATGGCTAA-3'	60.0	NIPBL	162	35
Nip1R	5'-AATTGCTTTGTTCCGAGCAT-3'	59.7	NIPBL	162	35
GAPDH1F	5'-GAGTCAACGGATTTGGTCGT-3'	60.0	GAPDH	265	25
GAPDH1R	5'-TTGATTTTGGAGGGATCTCG-3'	60.0	GAPDH	265	25

2.5 Preparation and handling of protein

2.5.1 Preparation of whole cell extracts

Whole cell protein extracts were prepared by trypsinising cells and centrifuging at 1200*g* for 4min. The cells were then washed in ice-cold PBS, and an appropriate volume of 2 X SDS loading buffer was added to lyse the cells. The solution was then transferred to a microcentrifuge tube and boiled at 95°C for 20min to denature the proteins. Samples were then sonicated briefly to shear the DNA. The samples were centrifuged at 16,000*g* for 10min at 4°C and the supernatant transferred to a clean tube.

2.5.2 SDS PAGE resolution of proteins

Protein extracts were resolved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Denaturing polyacrylamide mini-gels with 7-12% acrylamide (v/v), 0.39M Tris-HCl (pH 8.8), 0.1% SDS (w/v) 0.1% ammonium persulphate (w/v) and 0.04% *N*,*N*,*N'*,*N'*-tetramethyl-ethane-1,2-diamine (TEMED; v/v) were prepared in dH_2O and left to set with an isopropanol overlay. The alcohol was then removed and replaced with a stacking gel made of 5% acrylamide (v/v), 0.13M Tris-HCl (pH 8.8), 0.1% SDS (w/v) 0.1% ammonium persulphate (w/v) and 1% TEMED (v/v), topped with a comb. Once the gel was set, it was placed in an Invitrogen Novex Mini-Cell gel tank with SDS loading buffer, the samples loaded and run for ~2hrs at 110V. BenchMark™ and HiMark™ pre-stained protein standards from Invitrogen were run alongside samples to aid analysis.

Proteins were visualised by treating a gel overnight with GelCode Blue (Thermo Scientific) at 4°C with shaking. The stain was then removed by rinsing in tap water. The staining was used to estimate the approximate protein concentration, so that equal quantities of protein were used in western blotting.

2.5.3 Western blotting

After resolution on an SDS-PAGE gel, proteins were transferred to nitrocellulose membrane (Whatmann) using a Genie Blotter (Idea Scientific). The membrane and the gel were separately soaked in Western Transfer Buffer for 5min. The apparatus was assembled following the manufacturer's instructions sandwiching the gel and the membrane between 3mm paper (Whatmann), with the gel towards the cathode and the membrane towards the anode. The proteins were transferred for 1hr 30min at 24V, 2.5A.

Membranes were first blocked for 1hr at room temperature in TBST (TBS with 0.05% Tween 20) with 3% BSA to reduce background. Primary antibodies were then diluted to an appropriate concentration (Table 2.3) in 1% BSA in TBST, and were incubated with the membrane for 1hr at room temperature. The membrane was then washed in TBST three times for 10min each time, and was then incubated for 1hr with horseradish peroxidase conjugated secondary antibodies, diluted to an appropriate concentration (Table 2.3) in 1% BSA in TBST. The membrane was washed again, as above.

Membranes were developed using enhanced chemiluminescent (ECL) detection with SuperSignal Western Pico Reagent from Pierce. Signals were exposed on Hyperfilm Ecl (Amersham-Pharmacia).

2.6 Cell culture

2.6.1 Cell lines

Ten human Epstein-Barr virus- (EBV-) transformed lymphoblastoid cell lines (LCLs) were used, four control lines from healthy probands and six with NIPBL mutations (Table 2.4), obtained from Professor Tom Strachan at the University of Newcastle, and from Dr. Matt Deardorff at the Children's Hospital of Philadelphia. The HT1080, a human fibrosarcoma cell line (Rasheed *et al*, 1974), was used for RNAi and transfection studies.

Table 2.3 - Antibodies used

Antibody	Supplier	Catalogue Number	Species	Stock concentration (mg/ml)	Western Blot dilution	Immunofluorescence for FACS dilution	Immunofluroescence dilution	FISH Dilution
Primaries								
CTCF (Human)	Upstate	07-729	Rabbit	unknown	1/1000	1/800		
SMC1 (Human)	Bethyl	A300-055A	Rabbit	1	1/1000	1/800		
V5 Tag	Invitrogen	R960-25	Mouse	1.2	1/5000			
GAPDH (Human)	Abcam	ab9485	Rabbit	1	1/2000			
BrdU	BD Pharmingen	555627	Mouse	0.5			1/10	
FITC-anti-DIG	Roche	11 207 741 910	Sheep	1				1/20
TR-avidin	Vector	A-2016	Goat	2				1/500
FITC-avidin	Vector	A-2001	Goat	5				1/500
Rhodamine-anti-DIG	Roche	11 207 750 910	Sheep	1				1/20
Secondaries								
Anti-Rabbit IgG HRP	Sigma	A0545	Goat	unknown	1/16000			
Anti-Rabbit IgG FITC	Jackson	711-095-152	Donkey	0.75		1/1000		
Anti-Mouse IgG HRP	Sigma	A3682	Goat	unknown	1/5000			
Anti-Mouse IgG Texas Red	Jackson	715-075-150	Donkey	1.5			1/100	
FITC-anti-sheep	Vector	FI-6000	Rabbit	1.5				1/100
TR-anti-sheep	Vector	TI-6000	Rabbit	1.5				1/100
Biotin anti-avidin	Vector	BA-0300	Goat	0.5				1/100

Table 2.4 - CdLS LCLs used

CdLS or control	Cell line name	Original name	Mutation type	Gene mutation	Amino acid change	CdLS patient phenotype	From	Address
Control	W1	575						
	WP1	GIA 001 S					Matt Deardorff	Children's hospital of Philadelphia, USA
	WP2	CdL 043 S1					Matt Deardorff	
	WP3	CdL 047 S					Matt Deardorff	
CdLS	CP1	CdL 125 P	Frameshift	2479_2480 del AG	R827G fsX2	Severe	Matt Deardorff	
	CP2	CdL 186 P	Frameshift	226_230 del CACAT	H76R fsX5	Severe	Matt Deardorff	
	CP3	CdL 201 P	Nonsense	1604 C>G	S535X	Severe	Matt Deardorff	
	CP4	CdL 223 P	Microdeletion	del exons 2-17	No start codon	Severe	Matt Deardorff	
	CN1	AG0805	Splice site	5575-2 A>G	delete exon 30	Severe	Tom Strachan	Institute of Human Genetics, Newcastle, UK
	CN2	AG0088	Frameshift	7306_7307 ins G	S2435X	Severe	Tom Strachan	

2.6.2 Storing cells in liquid nitrogen

Approximately 2-5 X 10^6 LCL cells were frozen in 1ml foetal calf serum (FCS) supplemented with 7% dimethyl sulphoxide (DMSO), at -80°C in a polystyrene box, then transferred to liquid nitrogen after one week. To freeze HT1080s, one t25cm² tissue culture flask was split between two aliquots of 1ml FCS with 7% DMSO, and frozen as above.

Cells were thawed at 37°C, then diluted in tissue culture media, and centrifuged at 1200*g* for 4min. The media was aspirated from the cells to remove the DMSO. LCLs were then suspended in 5ml media and seeded into a t25cm² flask and HT1080s were suspended in 15ml media and seeded into a t75cm² flask.

2.6.3 Culture of cell lines

LCLs were grown in Invitrogen Roswell Park Memorial Institute (RPMI) 1640 media, supplemented with 10% FCS, 1% Penicillin/Streptomycin and 1% L-Glutamine. Cells were grown at 37°C with 5% CO_2 in upright flasks. The volume of culture media was doubled by adding an equal volume of media when the cells had approximately 2-5 X 10^5 cells/ml as judged by eye (as LCLs form clumps it is possible to judge their approximate concentration without a microscope). Once the flask was one third full, the cells were split by centrifugation at 1200*g* for 4min, and the media taken off. They were flicked to dislodge the cells then resuspended in an equal volume of media as before splitting and split between two new flasks. LCLs are cultured at containment level 2, following local rules.

HT1080s were grown in Invitrogen Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% FCS, 1% Penicillin/Streptomycin and 1% L-Glutamine. Cells are grown at 37°C with 5% CO_2. Cells were grown to near confluency then split. The media was aspirated off, then the cells were washed with PBS, before addition of a small amount of 1:1 Trypsin-Versene for 1min at 37°C. The activity of the enzyme was stopped by addition of supplemented DMEM, and the cells pelleted at 1200*g* for 4min. The cell pellet was resuspended in fresh media and seeded into appropriate tissue culture flasks. Cells were split at between 1:2 to

1:4, depending on when the cells were required. HT1080 cells were cultured at containment level 1, following local rules.

2.6.4 Sorting cells by centrifugal elutriation

LCLs were separated into the three stages of interphase (G_1, S and G_2) by centrifugal elutriation (CE). The few cells in mitosis would separate out with G_2 cells. CE separates cells on the basis of size by centrifugation combined with a pump. A Beckman J6-MC elutriator was used combined with a Masterflex pump. 10^8 cells were suspended in 10ml Elutriation Buffer. These were pumped into the elutriation chamber with the centrifuge spinning at 2500rpm at 16°C, with the pump set to 50 (this is an arbitrary setting on the pump), an approximate rate of 1ml/s. Elutriation buffer was pumped through the system and the pump speed was increased to 60, and 100ml of the first fraction was taken off. 100ml of each fraction was taken, increasing the pump speed by 10 each time, until fraction 9 (pump speed 140), then fraction 10 was taken at pump speed 140, reducing the centrifuge to 2000rpm.

Each fraction was split in two, one half was fixed in MAA for FISH analysis (Section 2.8), and the other half prepared for fluorescence activated cell sorting (FACS) analysis (Section 2.9). Fractions were then analysed by FACS to determine the DNA content and hence the cell cycle stage. As cells become larger as they progress through the cell cycle, the fractions roughly corresponded to different cell cycle stages.

2.6.5 Pulsing cells with Bromodeoxyuridine to mark DNA replication

10ml of cell suspension was transferred to a Petri dish. 10µl of 10mM BrdU in PBS was added to the cells, giving a final concentration of 10µM BrdU. The cells were incubated at 37°C for two hours.

2.7 RNA interference

2.7.1 Preparation of stock solutions of siRNA

ON-TARGETplus siRNA pools were purchased from Dharmacon (Table 2.5).

siRNA was supplied as a 5nmol pellet in a microcentrifuge tube, which was briefly centrifuged, and resuspended in 250µl siRNA buffer by shaking for 30min, giving a 20µM stock. The siRNA stock was aliquoted and frozen at -20°C.

2.7.2 Transfection of cells with siRNA pools

HT1080 cells were diluted to a plating density of 5 X 10^4 cells/ml in antibiotic-free media, and 2ml were placed in wells of 6-well plates. The cells were grown overnight at 37°C in 5% CO_2.

The stock siRNA solution was diluted to give a 2µM solution. 100µl siRNA and 100µl serum-free medium were mixed in one tube and incubated for 5min at room temperature. In a separate tube 4µl Dharmafect 4 reagent and 196µl serum-free medium were mixed and incubated for 5min at room temperature. The contents of the two tubes were mixed together and incubated for 20min at room temperature. 1.6ml of medium with 5% FCS and no antibiotic was added to the tube, giving 2ml of transfection mixture containing 100nM siRNA. The medium was removed from the cells plated out the previous day, and replaced with the transfection mixture. The cells were then cultured at 37°C until needed.

Two controls were used in every RNAi experiment, a mock transfection and a non-coding negative control. In the mock transfection, siRNA buffer was used instead of siRNA. In the non-coding control, an siRNA pool was used that does not correspond to any known genes.

RNAi treated cells were harvested after two days. Knockdown of gene expression was demonstrated by Western Blotting or by RT-PCR. Cells could then be harvested for either FACS or FISH.

Table 2.5 - RNAi pools used

Gene	Target sequences
NIPBL	5'-GAUAUAAACCGCCCACUAA-3' 5'-CAACAGAUCACAUAGAGUU-3' 5'-ACACUUCACUUCUAACAAA-3' 5'-GGGCUUGUUUCAAUAGAUA-3'
SMC1	5'-GGACAGCUCUAUUUGAAGA-3' 5'-GCUCGUAACUUCCUCGUUU-3' 5'-GUACAAGGGUCGACAGAUU-3' 5'-GAACUGGCCUCAAAGAACA-3'
CTCF	5'-GAUGAAGACUGAAGUAAUG-3' 5'-GGAGAAACGAAGAAGAGUA-3' 5'-GAAGAUGCCUGCCACUUAC-3' 5'-GAACAGCCCAUAAACAUAG-3'
Non-coding	unknown

2.8 Fluorescence *in situ* hybridisation (FISH)

2.8.1 Preparation of FISH probes

Probes were labelled with biotin-16-dUTP (bio-16-dUTP) and Digoxigenin-11-dUTP (DIG-11-dUTP; Roche) by nick translation. 4µl nick translation salts, 5µl of 0.5mM each of dATP, dCTP and dGTP, 5µl of 1mM bio-16-dUTP or 2µl of 1mM DIG-11-dUTP and 3µl of dTTP, 12µl DNA (~1µg), 2µl 1:500 DNase 1 (Invitrogen) and 2µl DNA polymerase 1 (Invitrogen) were combined and incubated at 16°C for 90min. The reaction was stopped by adding 55µl TE, 3µl 20% SDS and 2µ1 EDTA pH 8.0. Unincorporated nucleotides were removed by running through a Quick Spin Column (Pharmacia) following the manufacturer's instructions.

Chromosome paints were prepared by PCR. A 50µl mix was made up containing 200ng DOP primer (5'-CCGACTCGAGNNNNNNATGTGG-3'), 1 X PCR buffer, 25mM $MgCl_2$, 2mM of each dATP, dCTP and dGTP, 0.5mM dTTP, 1mM DIG-11-dUTP and 1µl template DNA (~500ng). A PCR was set up as below:

1. 94°C for 3min
2. 94°C for 45s
3. 60°C for 1min
4. 72°C for 1min, increasing by 5s for each cycle
5. Return to step two, 29 times
6. 72°C for 10min

Labelling was tested by running the labelled paints in a 2% agarose gel, a successful paint giving a smear at around 300bp. Unincorporated nucleotides were cleaned up by running through a Quick Spin Column (Pharmacia) following the manufacturer's instructions. Paints for chromosomes 11 and 13 were prepared by this method; paints for chromosomes 17 and 18 were purchased from StarFISH.

2.8.2 Quantification of label incorporation

Labelling was detected using either Streptavidin alkaline phosphatase or anti-DIG alkaline phosphatase. Gridded nitrocellulose filters were washed with dH_2O followed by 20 X SSC, then dried. 1/500, 1/1000, 1/5000 and 1/10,000 dilutions of

probes were spotted onto the filter, along with 20pg, 10pg, 2pg and 1pg labelled lambda DNA standards. The DNA was crosslinked onto the filter by 1500mJ UV irradiation in a Stratalinker. The filter was washed for 5min in 0.1M Tris pH 7.5, 0.15M NaCl and then incubated for 30min at 60^{o}C with this buffer containing 3% BSA. It was then incubated in the same buffer with 1% streptavidin alkaline phosphatase and/or anti-DIG alkaline phosphatase for 15min at room temperature. It was washed twice in this buffer alone for 15min then washed once in 0.1M Tris pH 9.5. The filter was then sealed in a polythene bag containing 5ml 0.1M Tris pH 9.5 with two drops of each of buffers 1, 2 and 3 from a Vector BCIP/NBT kit. The presence of alkaline phosphatase gives a blue colour due to the reaction of 5-bromo-4-chloro-3-idolyl-phosphate and nitroblue tetrazolium. The membrane was incubated in the dark until blue spots appear which could be compared to standards of known concentration to estimate the incorporation of labelling of each probe.

2.8.3 Fixing cells for FISH

2.8.3.1 Harvesting and fixing cells in 3:1 methanol:acetic acid for 2D FISH

20ml of LCLs at ~2-5 X 10^5 cells/ml, or one well of a 6-well plate of HT1080s was harvested then washed once in PBS before treatment with ~10ml hypotonic solution of 33mM KCl and 17mM tri-sodium citrate for 10-15min to swell the cells. They were centrifuged at 1200*g* for 4min and the hypotonic solution taken off. The cells were then fixed with 3:1 MAA for one hour; the cells were resuspended by adding MAA to them drop by drop whilst vortexing them, to prevent them clumping together. The cells were then centrifuged and put through three more washes in MAA of 5min each. MAA fixed nuclei were stored at –20^{o}C until needed.

Glass slides were stored in dilute solutions of HCl in EtOH, and were dried and polished with muslin prior to use. The fixed cells were spun down again at 1200*g* for 4min, and resuspended in a few drops of MAA to give a milky consistency. A slide was moistened by breathing onto it then single drop of this suspension was dropped from a height of around 20cm onto the slide, and blown upon. The quality of the

spread could then be monitored by phase contrast microscopy. Slides were then aged for 24 hours or more at room temperature.

2.8.3.2 Cytospinning and fixing cells for 3D FISH

Lymphoblastoid cells were prepared for 3D FISH by cytospin. 500µl confluent LCLs were harvested and resuspended in 10ml PBS. 500µl of the cell suspension was then applied to a Shandon Single Cytofunnel (Thermo Scientific), clipped onto a glass slide. This was centrifuged at 600rpm for 5min in a Shandon Cytospin 3, then the slide was removed and allowed to dry.

Cells were fixed to the slides using 4% paraformaldhyde (pFA) in PBS for 10min at room temperature, then washed three times in PBS. The slides were then placed in 0.5% Triton X-10 in PBS for 10min to permeablise the cells. They were washed again in PBS three times and stored at 4°C in PBS until needed.

2.8.4 Slide preparation for FISH

2.8.4.1 Slide preparation for 2D FISH

Slides were treated with 100µg/ml RNase in 2 X SSC at 37^{o}C for 1hr to digest any RNA in the cell preparation. Cells were dehydrated through a series of EtOH washes consisting of 2min in each of 70%, 90% then 100% EtOH at room temperature then air-dried. The slides were heated to 70 ^{o}C for 5min, then denatured in 70% formamide in 2 X SSC at pH 7.5 at 70^{o}C for 60-75s depending on the age of the slides, then washed in ice cold 70% EtOH and through 90% and 100% EtOH washes as before, then air-dried.

2.8.4.2 Slide preparation for 3D FISH

Slides were treated with 100µg/ml RNase in 2 X SSC at 37^{o}C for 1hr to digest any RNA in the cell preparation. Cells were dehydrated through a series of EtOH washes consisting of 2min in each of 70%, 90% then 100% EtOH at room temperature then air-dried. Slides were treated with 0.1M HCl at 37°C for 5min then

dehydrated through alcohols as before. The slides were then heated to 70 ^{o}C for 5min, then denatured in 70% formamide in 2 X SSC at pH 7.5 at 80^{o}C for 10min, then washed in ice cold 70% EtOH and through 90% and 100% EtOH washes as before, then air-dried.

2.8.5 Hybridisation of FISH probes

For interphase separation measurements 100ng of each of two fosmid probes were prepared with 5µg human Cot I (Invitrogen) and 0.5µg salmon sperm DNA. For CTs, 150ng of chromosome paint for each arm was prepared with 14µg human Cot I and 0.5µg salmon sperm DNA. For "looping out" studies, 150ng of chromosome paint for each arm and 100ng of probe was prepared with 16.5µg human Cot I and 0.5µg salmon sperm DNA. The DNA precipitated with two volumes of 100% EtOH then pelleted and dried out in a spin vac. The dried DNA was dissolved in hybridisation mix for one hour. The probe was denatured at 70^{o}C for 5min, reannealed at 37^{o}C for 15min then applied to the slides and sealed under a coverslip with TipTop rubber solution from Rema. The slides were incubated overnight in a covered tray floating in a 37°C waterbath to hybridise.

2.8.6 Washing and detection of FISH signals

The glue was peeled off then hybridisation mix was washed off the slides using four 3min washes in 2 X SSC at 45°C followed by the same in 0.1 X SSC at 60°C. To reduce background, the slides were blocked in 5% Marvel in 4 X SSC, prior to probe detection using a series of antibody conjugations. For interphase separation, DIG was detected using sequential layers of Fluorescein isothiocyanate (FITC)-conjugated anti-digoxygenin and FITC-conjugated anti-sheep IgG and biotin was detected with sequential layers of Texas Red (TR)-conjugated avidin, biotinylated anti-avidin (BAA) and TR-conjugated avidin again. For chromosome paints, DIG was detected in FITC-conjugated avidin, followed by BAA then FITC-conjugated avidin again and biotin detected in Rhodamine-conjugated anti-DIG followed by TR-conjugated anti-sheep. All antibodies were diluted to an appropriate concentration (Table 2.3) in 5% Marvel in 4 X SSC. The slides were incubated with the antibodies

for 30min at 37°C in a humidity chamber with washing three times in 4 X SSC with 1% Tween 20 for 2min after each conjugation.

2D FISH slides were mounted in Vectashield (Vector) containing 0.5µg/ml DAPI and sealed with PANG rubber solution. 3D FISH slides were incubated in 50ng/ml DAPI in 4 X SSC with 1% Tween for 5min before mounting in Vectashield (Vector) and sealed with PANG rubber solution.

2.8.7 Dectection of BrdU incorporation

To detect BrdU incorporation, cells were dropped onto slides and prepared as for 2D FISH (Section 2.8.3.1; Section 2.8.4.1). They were then blocked for 5min in 4 X SSC with 5% Marvel at room temperature. Two antibody conjugations were applied, first anti-BrdU, followed by Texas-red anti-mouse, both diluted in block solution, incubating for 30min in a humidity chamber at room temperature then washing three times for 2min in 4 X SSC with 1% Tween after each incubation. The slides were then mounted in Vectashield (Vector) containing 0.5µg/ml DAPI and sealed with PANG rubber solution (PANG).

2.8.8 Image capture

2.8.8.1 Imaging nuclei in 2D

2D nuclei were photographed using an imaging system which comprises a Coolsnap HQ CCD camera (Photometrics Ltd, Tucson, AZ) Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss, Welwyn Garden City, UK) and Chroma #83000 triple band pass filter set and Chroma #83700 emission filters (Chroma Technology Corp., Rockingham, VT). The single excitation and emission filters are installed in motorised filter wheels (Prior Scientific Instruments, Cambridge, UK). Image capture and analysis were performed using in-house scripts written for IPLab Spectrum (Scanalytics Corp, Fairfax, VA).

2.8.8.2 Imaging nuclei in 3D

3D nuclei were photographed using an imaging system comprising a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK), Zeiss Axioskop fluorescence microscope with Plan-neofluar or Plan apochromat objectives, a Lumen 200W metal halide light source (Prior Scientific Instruments, Cambridge, UK) and Chroma #83000 triple band pass filter set (Chroma Technology Corp., Rockingham, VT) with the excitation filters installed in a Prior motorised filter wheel. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Hardware control, image capture and analysis were performed using in-house scripts written for IPLab Spectrum (Scanalytics Corp, Fairfax, VA). Stacks of 20 images at 0.25μm intervals were taken.

2.9 FACS

2.9.1 Detection of the DNA content of cells

2.9.1.1 Fixing and staining LCLs

Cell suspensions were transferred to 5ml BD Falcon tubes and centrifuged at 1200*g* for 4mins. The medium was removed and the cells resuspended in 100μl citrate buffer. These could then be stored at -20°C until needed.

On the day of FACS analysis, 450μl FACS Solution A was added to cells, mixed and the cells incubated for 2min at room temperature. 375μl FACS Solution B was then added, and the cells incubated at room temperature for a further 10 min. The cells were then incubated in the dark on ice for 10min with 250μl FACS Solution C, and kept on ice in the dark until needed.

2.9.1.2 Fixing and staining HT080 cells

RNAi treated HT1080 cells were trypsinised and centrifuged at 1200*g* for 4min. The media was removed and the cells washed in ~10ml of cold PBS. The cells were spun down again and the PBS poured off. The tubes were flicked to resuspend the

cells in the residual PBS, around 100µl and 1ml of ice-cold 70% EtOH was added drop-wise while vortexing. The cells were incubated on ice for 30min, and were stored at -20°C until needed.

DNA was stained with Propidium Iodide (PI) by adding 50µl of 1mg/ml RNase and 50µl of 400µg/ml PI. The cells were incubated at 37°C for 30min in the dark and held in the dark on ice until FACS analysis.

2.9.2 Immunostaining for FACS analysis

To combine immunostaining with FACS analysis, RNAi-treated HT1080s were fixed as above, then immunostained.

For immunostaining cells were pelleted by centrifugation at 1200*g* for 4min and the cells washed in 2ml 5% donkey serum/PBS, then spun down again, resuspended in 100µl 5% donkey serum/PBS and incubated at room temperature for 10min. The primary antibody was added to the appropriate concentration (Table 2.2) and the cells incubated at room temperature for one hour. The cells were washed twice in 5% donkey serum/PBS, centrifuging at 2400*g* for 4min in between washes, then resuspended in 100µl 5% donkey serum/PBS with the secondary antibody diluted to an appropriate concentration (Table 2.2). The cells were incubated for 30min in the dark, then washed in 5% donkey serum/PBS, followed by PBS only, centrifuging at 2400*g* for 4min in between washes and resuspended in 500µl PBS.

DNA was stained with Propidium Iodide (PI) by adding 50µl of 1mg/ml RNase and 50µl of 400µg/ml PI. The cells were incubated at 37°C for 30min in the dark and held in the dark on ice until FACS analysis.

2.9.3 FACS analysis

Flow cytometric analysis was performed using a BD FACSAriaII SORP (Becton Dickinson). A 488nm laser was used to measure PI fluorescence (685/35nm bandpass filter) and FITC fluorescence (525/50nm bandpass filter). BD FACSDiva software (Becton Dickinson, Version 6.1.2) was used for instrument control and data

analysis. Further data analysis was carried out using FlowJo (Treestar Inc., Version 7.5.5) software.

2.10 Bioinformatics

2.10.1 Comparison of published microarray expression data

The datasets compared by bioinformatics are shown in Table 2.6. The gene-names and affymetrix probe symbols in the non-human datasets were converted to the names of their human orthologues using Biomart (http://www.biomart.org/). The data was then compiled into a single table in Microsoft Excel of the $\log_2$ fold change of each gene. Microsoft Excel software was used to compare correlation between datasets.

2.11 Data analysis

2.11.1 General statistical and quantitative analysis

Nuclear area datasets were analysed using Mann-Whitney U-tests (MW tests) in MiniTab 15 software. The MW-test is a non-parametric test, so the datasets can follow any distribution (unlike the Student's *t*-test, which assumes all datasets follow a normal distribution), however it assumes that both datasets tested follow a similar, but shifted, distribution. It was chosen because nuclear area datasets tend to follow a Rayleigh distribution, with a distinct cut-off at zero at one end of the distribution, and tending towards infinity at the other end of the distribution. Two datasets are assumed to be significantly different if $p \leq 0.05$.

Large FISH datasets were first analysed using Analysis of Variance (ANOVA). As FISH datasets give a Rayleigh distribution and ANOVA requires data to follow Normal distributions, the $\log_{10}$ of each value was first calculated, then ANOVA carried out on the group to determine if there was any significant difference between all members of the group. The group was considered to be significantly varied if $p \leq 0.05$. A Tukey's range test was carried out pairwise on the $\log_{10}$ datasets to give confidence intervals for each, a pair was considered to be significantly different if the

Table 2.6 - Published datasets used for bioinformatics

Species	Cell line	Method of CdLS modelling	Published in	Table in paper
Human	LCLs	CdLS patients	Liu *et al, 2009*	S2 and S3
Mouse	Brain	Nipbl gene-trap	Kawauchi *et al* , 2009	S4
Mouse	MEFs	Nipbl gene-trap	Kawauchi *et al* , 2009	S3
Mouse	ES	Nipbl RNAi	Kagey *et al* , 2010	S3
Mouse	ES	Smc1 RNAi	Kagey *et al* , 2010	S3
Drosophila	BG3	Nipped-B RNAi	Schaaf *et al* , 2009	S4

confidence intervals did not overlap. Pairwise comparisons were then made using MW tests on the raw data, as with the nuclear area data. All of these analyses were carried out using MiniTab 15.

FACS datasets were analysed using χ^2 tests, as these are much smaller datasets. χ^2 is calculated as:

$$\chi^2 = \Sigma \frac{(\text{Observed-Expected})^2}{\text{Expected}}$$

The value of χ^2 was then used to calculate the p-value using the P-value Calculator available online from GraphPad Software (http://www.graphpad.com/quickcalcs/pvalue1.cfm), using two degrees of freedom, as there are three cell cycle stages.

2.11.2 Analysis of fluorescence microscopy images

2.11.2.1 Analysis of nuclear area

To study nuclear area, fifty nuclei were selected which were evenly shaped and were not touching any adjacent nuclei and imaged in a 100 X objective in bin 2. The area of DAPI staining was measured for each nucleus in pixels, this was converted to microns by multiplying by $(0.134\mu m)^2$.

In the cell cycle nuclear area studies, cells in S-phase and G_2 were stained with BrdU, and were visualised in red, whilst cells in G_1 were not stained. Fifty nuclei of each type were selected and analysed by measuring the area of DAPI staining as above.

2.11.2.2 Analysis of chromosome territory area

For CT area studies, nuclei were selected that were evenly shaped and that were not touching another nucleus and imaged in a 100 X objective in bin 2. The chromosome territories had to be bright, easily visible, distinct from one another and there had to be only two of them.

An IPLab script written by Paul Perry was used to measure the CT area (Figure 2.1; adapted from Mahy *et al*, 2002a). The script first measured the area of DAPI staining, giving an estimate of nuclear area. Each chromosome was then identified using a circle tool and the FITC signal was detected within that area. The area of FITC signal above a threshold was then measured. This threshold was manually set each time. For analysis this area was normalised to the nuclear area.

2.11.2.3 Analysis of looping out from chromosome territories

For CT looping studies, nuclei were selected that were evenly shaped and that were not touching another nucleus and imaged in a 100 X objective in bin 2. The chromosome territories had to be bright, easily visible, distinct from one another and there had to be only two of them. The probe spots had to be bright and easily visible, and easy to determine which spot corresponded to which territory.

An IPLab script written by Paul Perry was used to measure the CT area (Figure 2.2; Mahy *et al*, 2002a). The script first measured the area of DAPI staining, giving an estimate of nuclear area. The chromosomes and their corresponding probes were then identified using a circle tool and a paint tool respectively. The FITC signal was detected within the circle tool area and the probe signal was detected within the paint tool area. An expanding circle from the probe signal was used to find the closest CT edge based upon a FITC intensity threshold, manually set each time, and the distance measured in pixels. For analysis this was normalised to the nuclear radius, estimated by taking the square root of the area divided by pi.

2.11.2.4 Analysis of interphase chromatin compaction in 2D

For interphase separation studies in 2D, nuclei were selected that were evenly shaped and that were not touching another nucleus and imaged in a 100 X objective in bin 2. The probe spots had to be bright and easily visible, there had to be only two pairs of the probe spots, as cells had to be diploid, the probe spots had to be single round spots rather than chains and the pairs had to be far enough apart to be resolved as individual pairs.

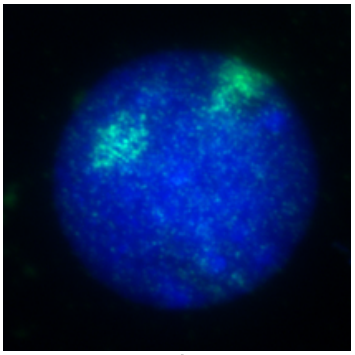
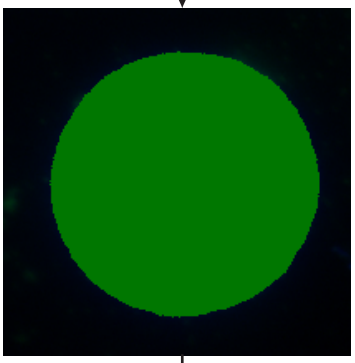
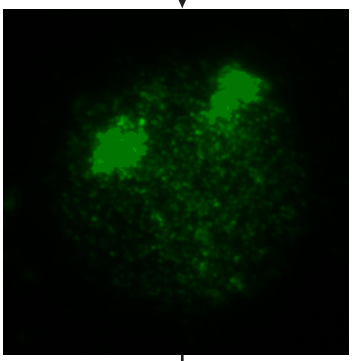


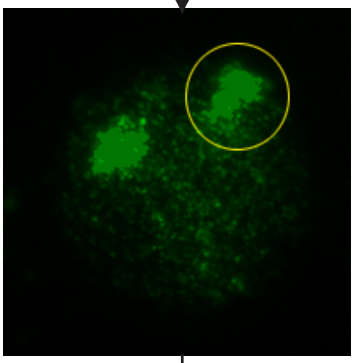
Image of a nucleus labelled with a chromosome paint



The image was segmented on the basis of DAPI staining intensity. The area of the segment was measured giving an estimate of the nuclear area



The image was split into the green colour channel and segmented on the basis of FITC staining intensity

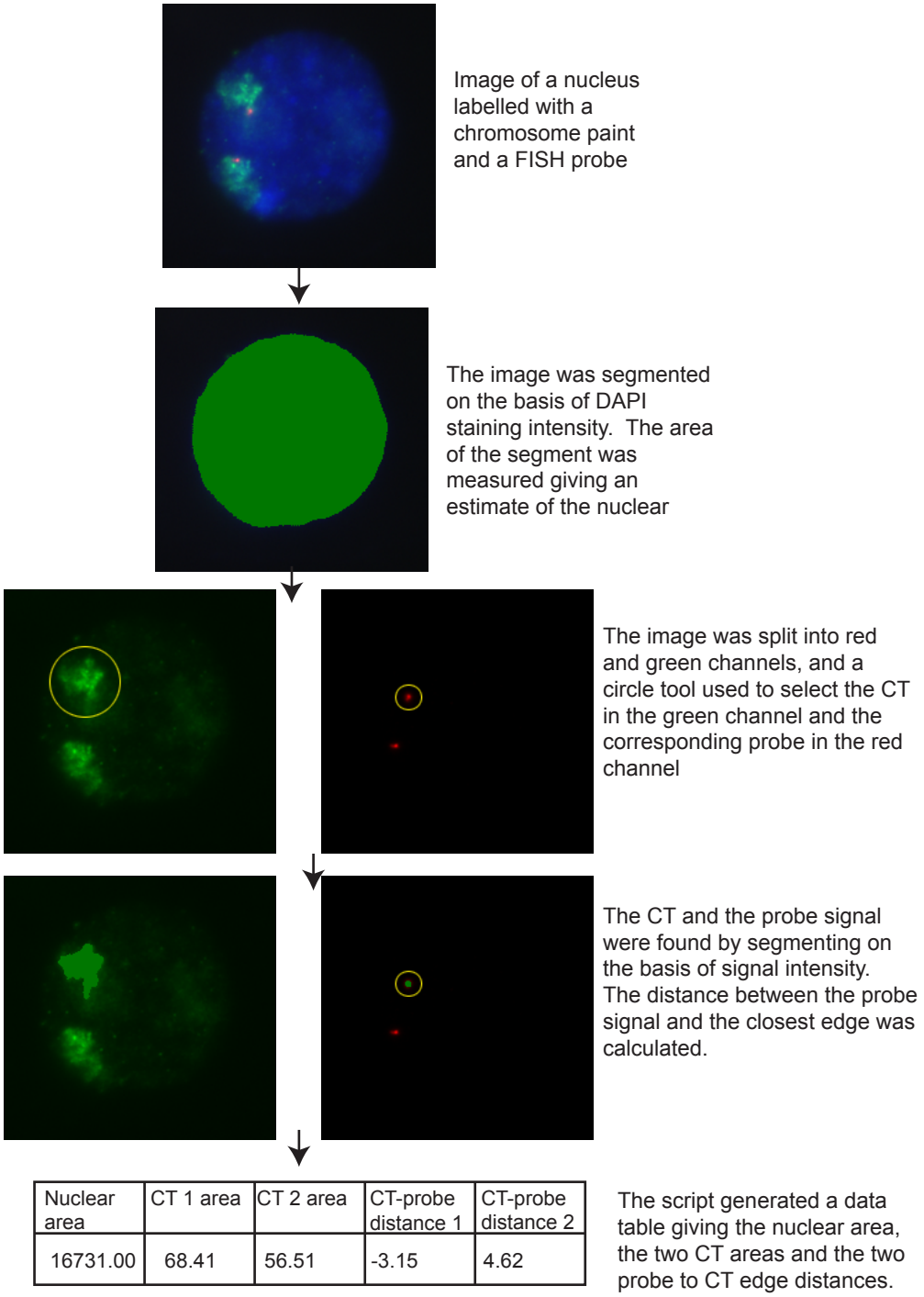


The chromosome territories were separately selected using a circle tool, and the area of FITC staining was measured in each one

Nuclear area	CT 1	CT 2
17651.00	54.93	63.81

The script generated a data table giving the nuclear area and the two CT areas.

Figure 2.1 – Measuring CT areas from images.
An illustration demonstrating the IPLab script, written by Paul Perry, which was used to measure the size of CTs in FISH stained nuclei.



Nuclear area	CT 1 area	CT 2 area	CT-probe distance 1	CT-probe distance 2
16731.00	68.41	56.51	-3.15	4.62

Figure 2.2 – Measuring the distance of probes from the CT edge from images.
An illustration demonstrating the IPLab script, written by Paul Perry, which was used to measure the distance of probes from the edge of CTs in FISH stained stained nuclei.

An IPLab script written by Paul Perry was used to measure the distance between the two probes of a pair (Figure 2.3; Eskeland *et al*, 2010). The script first measured the area of DAPI staining, giving an estimate of nuclear area. Pairs of probes from the same chromosome were identified using a paint tool; the script then identified the spots within the area and defined the centre. The distance between the centres of the two spots was measured in pixels, which could then be converted into microns by multiplying by 0.134μm, or normalised to nuclear area.

2.11.2.5 Analysis of interphase chromatin compaction in 3D

In 3D, stacks of 20 images at 0.25μm intervals were taken. Nuclei were selected on the same basis as 2D nuclei. The stacks were resolved into a single image using IPLab, taking the brightest signal at each spot. This image was then analysed as with the 2D nuclei, giving the distance between the probes in the *xy* plane. The distance in the *z* plane was calculated by manually examining the individual images in the stacks, and determining in which image the probe was in focus. The distance between pairs of probes in the *z* plane could then be calculated by counting the number of stacks in between the two images, in which the probes were most in focus, and multiplying the number of stacks by 0.25μm. The actual distance between the two probes could then be calculated from the *xy* distance (converted to microns by multiplying by 0.134μm) and the *z* distance, by Pythagoras' Theorum.

2.11.3 Analysis of FACS data

2.11.3.1 Analysis of cell cycle stages from FACS data

Cell cycle analysis of FACS data was carried out using FlowJo (Treestar Inc., Version 7.5.5) software. A DNA content distribution was calculated from the fluorescence pulse area (FL2-A) of the 685/35nm PI signal. A Watson pragmatic model was used to fit the DNA content distribution to the three cell cycle stages. In this model, the G_1 and G_2/M peaks are each assumed to follow a normal distribution with a mean at 2n and 4n chromosomes respectively. No assumptions are made

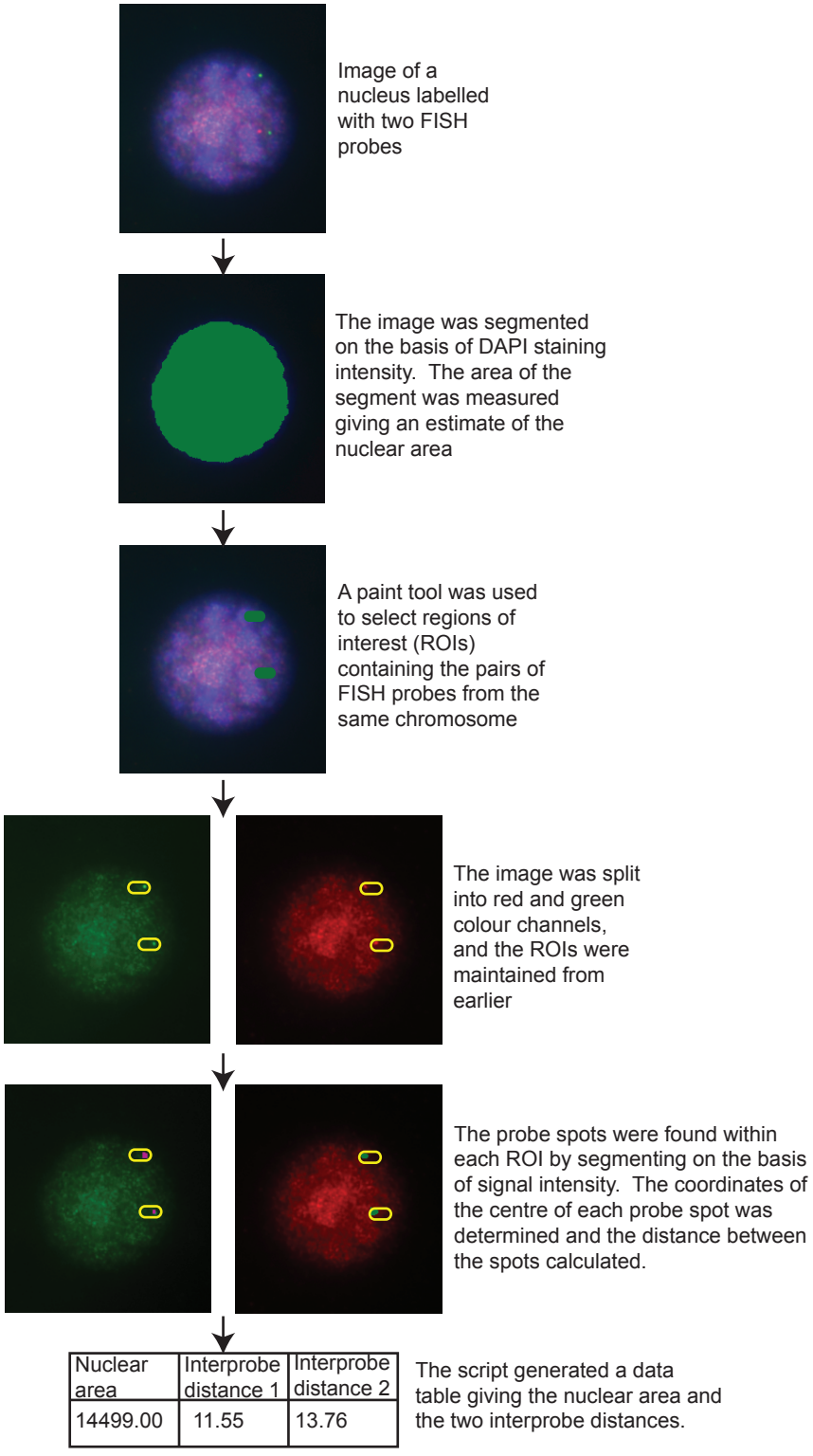


Figure 2.3 – Measuring interphase separation of probes from images.
An illustration demonstrating the IPLab script, written by Paul Perry, which was used to measure the physical distance between probe pairs in fluorescent images of FISH stained nuclei.

about the shape of the S-phase distribution, and the Watson model fits the distribution exactly (http://www.flowjo.com/v76/en/ccmodels.html).

2.11.3.2 Analysis of nuclear size from FACS data

Nuclear size was modelled from the fluorescence pulse width (FL2-W) of the 685/35nm PI signal. Flowjo (Treestar Inc., Version 7.5.5) software was used to calculate the mean fluorescence pulse width.

2.11.3.3 Analysis of immunostaining from FACS data

Immunostaining was measured from the fluorescence pulse area (FL2-A) of the 525/50nm FITC signal using Flowjo (Treestar Inc., Version 7.5.5) software. The FITC signal from an unstained control sample was used to set a gate for negative staining. This was then used to find the percentage of cells in each sample that scored as negative for FITC staining. The mean level of FITC staining was also calculated.

3 General effects of CdLS on the nucleus

In this project I wished to discover if, and how, chromatin is altered in CdLS. The fly homologue of NIPBL, Nipped-B, is involved in the interactions between promoters and enhancers (Rollins *et al*, 1999), and if NIPBL functions the same way, this could be the cause of the developmental phenotype of CdLS patients. I considered that if interactions between promoters and enhancers were physical interactions of chromatin, then disruption of a large number of these interactions might result in a gross physical change in the organisation of chromatin in CdLS cells compared to wildtype. In mouse ES cells, a large increase in nuclear area is observed upon RNAi knockdown of *Smc2*, a component of condensin, without any change in ploidy (Fazzio and Panning, 2010). As *NIPBL*, which is mutated in all the CdLS cell lines I studied, also loads condensin onto chromatin (D'Ambrosio *et al*, 2008), it seemed plausible that CdLS might cause a similar effect on nuclear size.

The first step was to use human CdLS cell lines to look for gross effects on the nucleus compared to wildtype cell lines. To this end I have studied nuclear area of CdLS and wildtype cells, as a possible measure of gross chromatin compaction of cells. It has been shown that addition of HeLa nuclei to Xenopus oocytes resulted in a massive increase in the size of the HeLa nucleus, far greater than expected from increase in nuclear material suggesting that oocytes contain an agent capable of altering HeLa nuclei (Gurdon, 1976). Addition of certain proteins to nuclei can result in a massive compaction of chromatin, giving smaller nuclei (Grigoryev *et al*, 1992), whereas deletion of the linker histone, H1, results in an increase in nuclear size (Eskeland *et al*, 2010). Therefore, measurement of nuclear size appears to be one way to measure chromatin compaction.

Due to the important role of cohesin in the cell cycle I also analysed the cell cycle of the same cells by FACS, determining the proportion of cells in each cell cycle stage. Finally, I used BrdU incorporation to determine cell cycle stage in fixed cells and measure nuclear area at different stages.

3.1 Some CdLS cell lines have increased nuclear area compared to wildtype

3.1.1 Preliminary data suggests increased nuclear size in CdLS

When I initially started this project in 2007, I had only two CdLS LCLs, AG0805 and AG0805, which had been kindly sent to us by Professor Tom Strachan at the University of Newcastle, named CN1 and CN2 in this thesis, respectively (Tonkin *et al*, 2004). CN1 has a splice site mutation that causes exon 30 of *NIPBL* to be omitted, however the coding sequence would remain in frame, so that a slightly shorter protein is predicted to be produced that lacks part of the HEAT repeats domain. CN2 has a frameshift mutation which results in a stop codon in exon 42, so encodes a C-terminally truncated protein. Both patients, from whom the cell lines were derived, were reported to have severe CdLS. I had a single control to compare these cell lines to, the commercially available 575 lymphoblastoid cell line, named W1 in this thesis (Figure 3.1; Table 2.3).

I decided to look for changes in the nuclear size, as this would reflect either a gross compaction or decompaction of chromatin. I fixed the three cell types in 3:1 methanol:acetic acid (Section 2.8.3.1), then dropped them onto slides and stained them with DAPI. The nuclei were visualised using fluorescence microscopy, and 100 images of each cell type were captured. I was then able to analyse the nuclear area of the cells using an IPLab script (Section 2.13.2.1) and used Mann-Whitney U-tests to compare the data from the three cell lines (Section 2.13.1).

In both CdLS cell lines studied, there was an increase in the average nuclear area compared to wildtype of approximately 38% ($p<0.0001$; Figure 3.1), the mean nuclear area of W1 cells being 257.7μm^2, and that of CN1 and CN2 being 377.1μm^2 and 336.7μm^2 respectively. However there was also a significant difference between the nuclear areas of the two CdLS lines ($p<0.001$). I thought that this increase in area might be due to a loss of cohesin-mediated intrachromosomal interactions, resulting in a general decompaction of chromatin. It was interesting to note that the

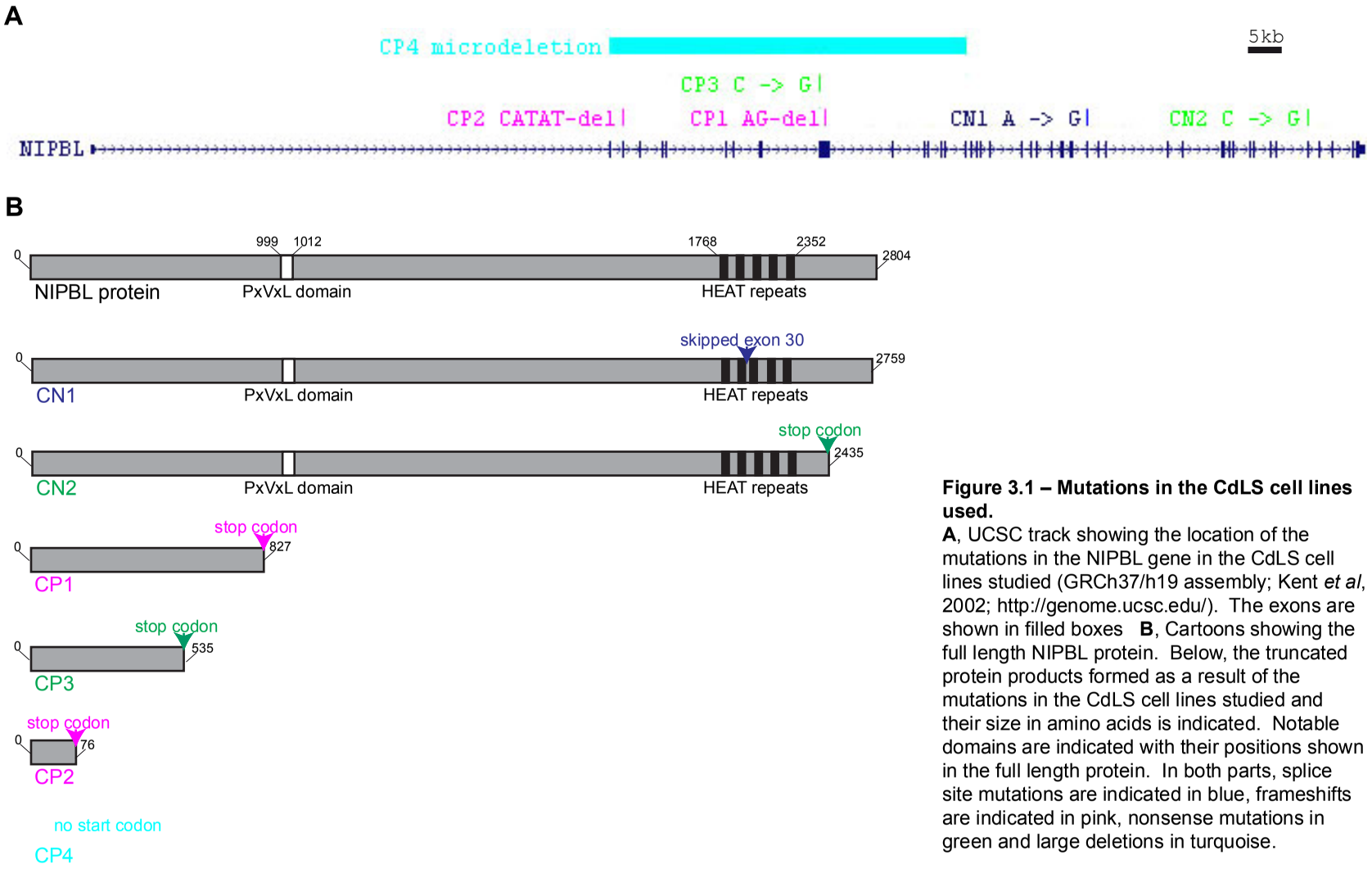


Figure 3.1 – Mutations in the CdLS cell lines used.
A, UCSC track showing the location of the mutations in the NIPBL gene in the CdLS cell lines studied (GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). The exons are shown in filled boxes **B**, Cartoons showing the full length NIPBL protein. Below, the truncated protein products formed as a result of the mutations in the CdLS cell lines studied and their size in amino acids is indicated. Notable domains are indicated with their positions shown in the full length protein. In both parts, splice site mutations are indicated in blue, frameshifts are indicated in pink, nonsense mutations in green and large deletions in turquoise.

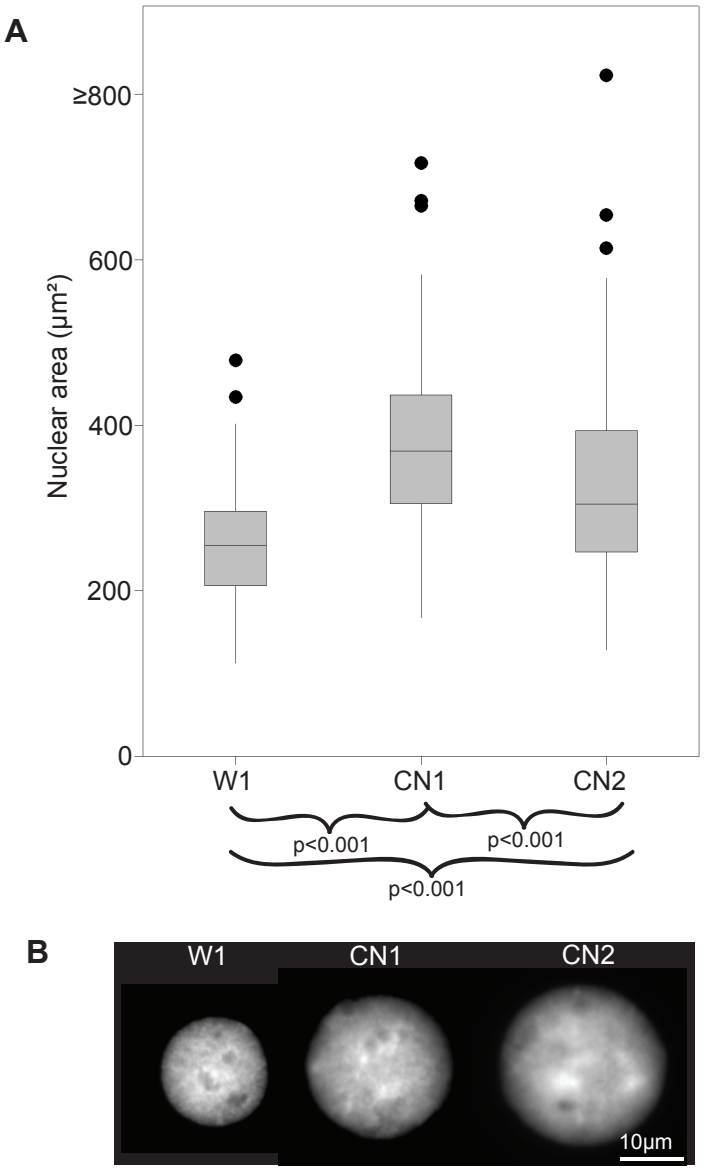


Figure 3.2 – The nuclear area of some CdLS cell lines is increased compared to wildtype.
A, Boxplot showing the distribution of nuclear area measurements (µm²; n=100) measured from wildtype LCLs (W1) and two CdLS LCLs (CN1 and CN2), including p-values from Mann-Whitney U-tests. The horizontal line represents the median, the boxes the interquartile range, the vertical lines the upper and lower limits and the dots the outliers. The upper limit is calculated as the upper quartile value plus 1.5 times the difference between the two quartiles, whilst the lower limit is the lower quartile minus 1.5 times the difference between the two quartiles. **B**, Sample images of DAPI stained nuclei from wildtype and CdLS LCLs.

effect in CN1 was greater than that in CN2, since mutation in the latter seems to have a larger effect on the predicted NIPBL protein (Figure 3.1)

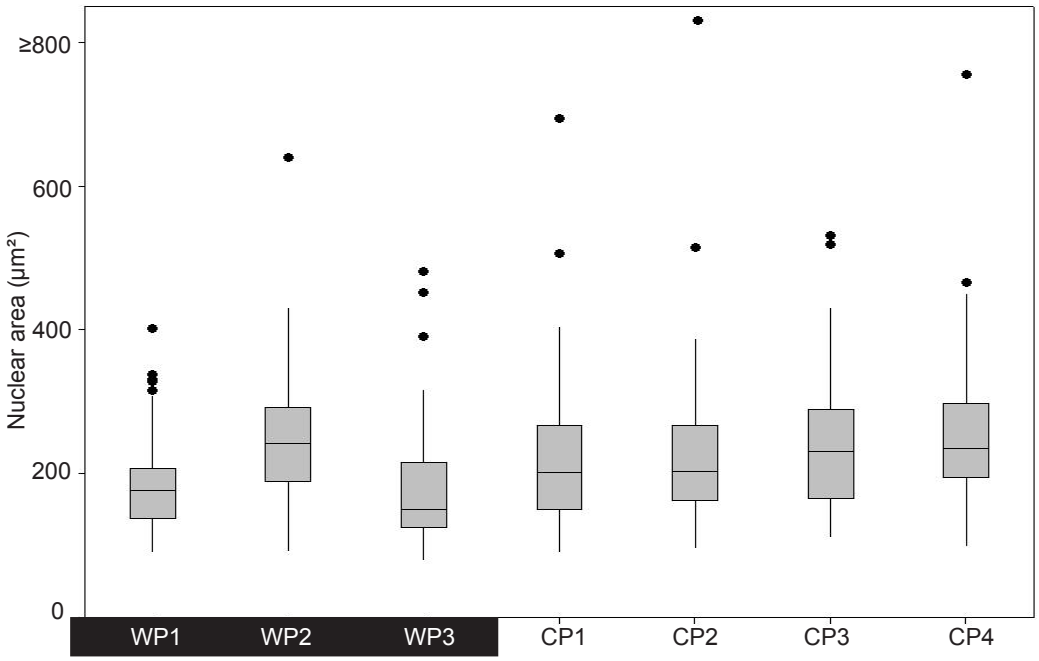
3.1.2 CdLS has a smaller effect on nuclear size in other cell lines

Later in my project I received a new set of lymphoblastoid cell lines from Dr. Matt Deardorff at the Children's Hospital of Philadelphia. These included three wildtype controls GIA 001 S, CdL 043 S1 and CdL 047 S, and four CdLS lines CdL 125 P, CdL 186 P, CdL 201 P and CdL 223 P; these cell lines were renamed WP1, WP2, WP3, CP1, CP2, CP3 and CP4 respectively for this thesis. The mutations in the CdLS cell lines were much more severe than in the CN1 and CN2 cell lines previously used. CP1 and CP2 both had frameshift mutations that resulted in early stop codons in *NIPBL*, CP3 had an early nonsense mutation and CP4 had a deletion of exons 2-17, including the start codon, thus these mutations are likely to result in either no protein at all, or severely truncated proteins lacking the PxVxL motif or the HEAT repeats (Figure 3.1; Table 2.3). All patients were reported to have severe CdLS.

I wanted to see if these new cell lines shared the nuclear phenotype found in the Newcastle cell lines. I carried out an identical analysis on the new cells (Figure 3.3).

The mean nuclear areas of the wildtype cells were WP1 (189.1μm^2), WP2 (257.3μm^2), and WP3 (204.1μm^2) giving a mean wildtype nuclear area of 216.8μm^2. The mean nuclear area of CdLS cells was 11% higher, at 241.2μm^2, and the means for the individual cell lines were CP1 (237.7μm^2), CP2 (222.7μm^2), CP3 (253.4μm^2) and CP4 (251.1μm^2). Note that the actual areas in this second set of cell lines are much lower than in the first set, however this does not reflect a biological difference between the two groups, as nuclear area of MAA fixed cells can be affected by the humidity on the day of the experiment; to control for this, each group of experiments were carried out simultaneously. Pair-wise Mann-Whitney U-tests were carried out on the datasets and I found that almost all cell lines, CdLS and wildtype have significantly different distributions of nuclear area compared to each other (Figure 3.3B). This is consistent with a general level of chromatin decompaction caused by *NIPBL* mutation.

A



B

	WP1	WP2	WP3	CP1	CP2	CP3	CP4
WP1	1	0	0.0192	0.0008	0	0	0
WP2	0	1	0	0.0004	0.0004	0.1131	0.9698
WP3	0.0192	0	1	0	0	0	0
CP1	0.0008	0.0004	0	1	0.7969	0.0399	0.0004
CP2	0	0.0004	0	0.7969	1	0.0577	0.0003
CP3	0	0.1131	0	0.0399	0.0577	1	0.1295
CP4	0	0.9698	0	0.0004	0.0003	0.1295	1

C

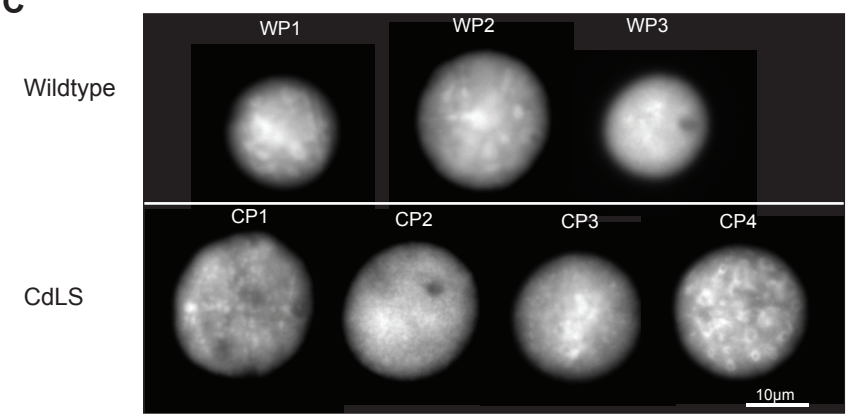


Figure 3.3 – The nuclear area of CdLS cell lines is increased compared to most wildtype cells. **A**, Boxplot showing the distribution of nuclear areas (µm²) in wildtype and CdLS cells (n>100), as **Figure 3.2**. **B**, p-values obtained from Mann-Whitney U-tests, comparing cell lines pair-wise. Wildtype cell lines are highlighted in black and CdLS cell lines are not highlighted. P-values ≤0.05 that denote a significant difference are highlighted in grey. **C**, sample images of nuclei stained with DAPI.

3.2 Cell cycle analysis in CdLS cell lines

I considered that CdLS cells might have an altered cell cycle compared to wildtype, due to the role of cohesin in the separation of chromatids at metaphase. It was important to exclude an effect on cell cycle as a possible explanation for the increased nuclear size in CdLS because increasing the DNA content through the cell cycle increases the nuclear size (Kang *et al*, 2010).

Therefore, I carried out FACS analysis on W1, CN1 and CN2, using PI staining as a measure of DNA content (Section 2.9; Figure 3.4). I firstly noticed that there was no evidence for increased ploidy in CdLS cell lines, which would be indicated by extra peaks of DNA content to the right of the G_2 peak. I used cell cycle gating tools on FlowJo software to determine the percentage of cells in each stage of the cell cycle, based on their DNA content (Section 2.12.3). Both CdLS cell lines had fewer cells in S-phase, and more cells in G_1 and G_2 than the W1 control cell line (Figure 3.4B). I carried out χ^2 statistical tests (Section 2.13.1), comparing the two CdLS cell lines to W1 and found that both were significantly different to wildtype ($p<0.001$).

When the new cell lines, WP1, WP2, WP3, CP1, CP2, CP3 and CP4, arrived, I also analysed them by FACS (Figure 3.5). The percentage of cells in each cell cycle stage was again calculated, and I carried out pair-wise χ^2 tests, comparing every cell line to every other cell line. I found that no cell line had a significantly different cell cycle distribution to any other cell line (Figure 3.5B). I carried out more χ^2 tests, comparing each of the Philadelphia cell lines to the three Newcastle cell lines, and found that the cell cycle of W1 did not significantly differ from any of the new cell lines, whereas both CN1 and CN2 had significantly different cell cycles from all the new cell lines, both wildtype and CdLS ($p<0.001$). I concluded that there is no consistent alteration in the cell cycle in CdLS cell lines. It is possible that CN1 and CN2 behave differently to the other CdLS cell lines because they were prepared in a different lab (Newcastle), and subtle differences in the EBV transformation procedure or cell culture conditions may have resulted in an altered cell cycle compared to LCLs prepared by other labs.

The efficiency of PI staining appeared to differ between the seven cell lines, with the histogram shifting along the x-axis for different cell lines, however the shape of

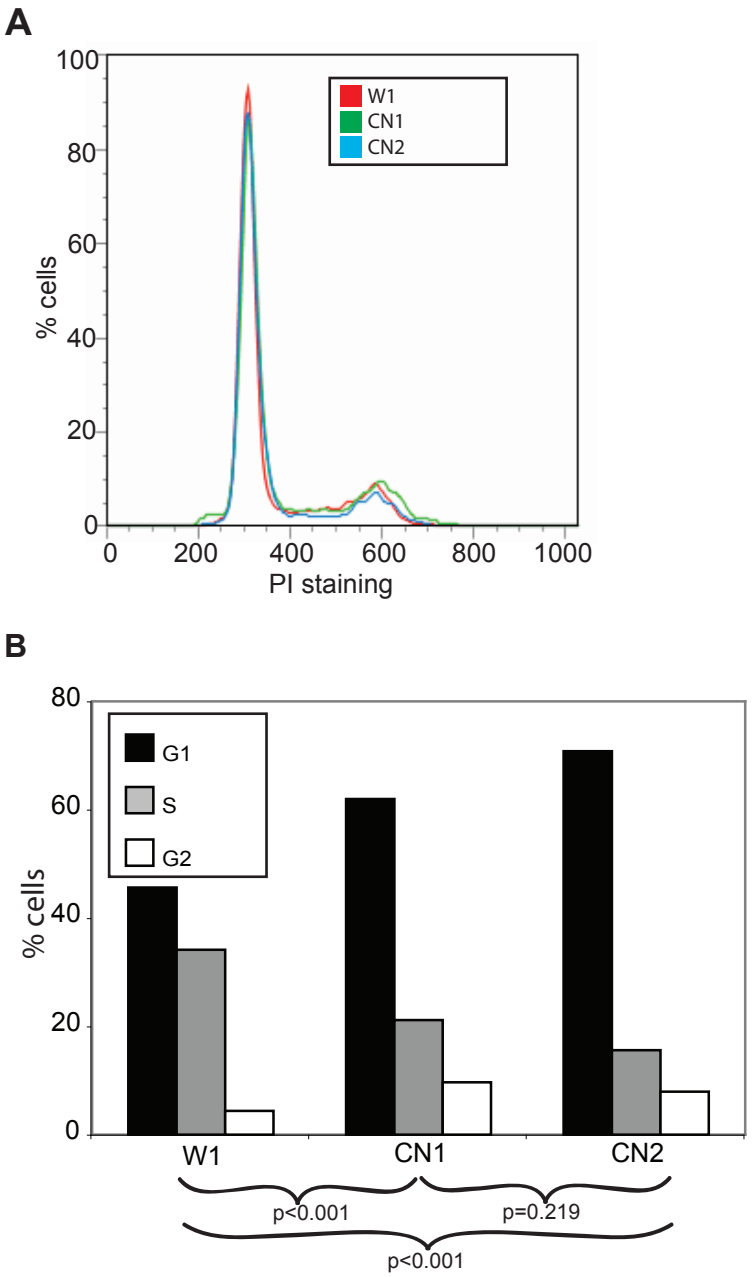
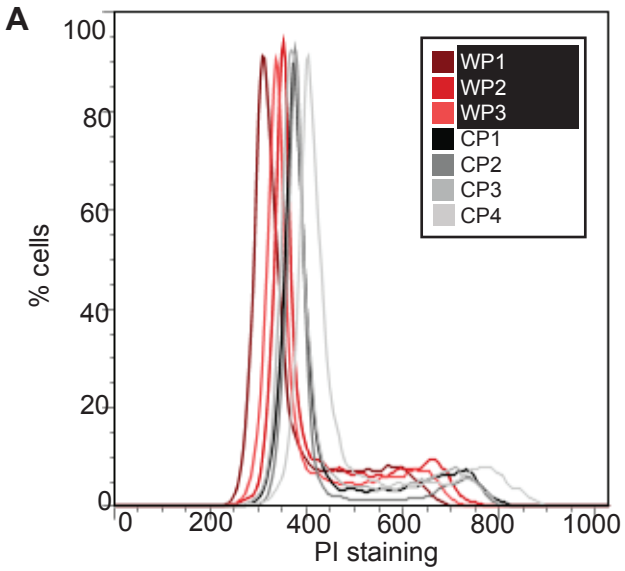
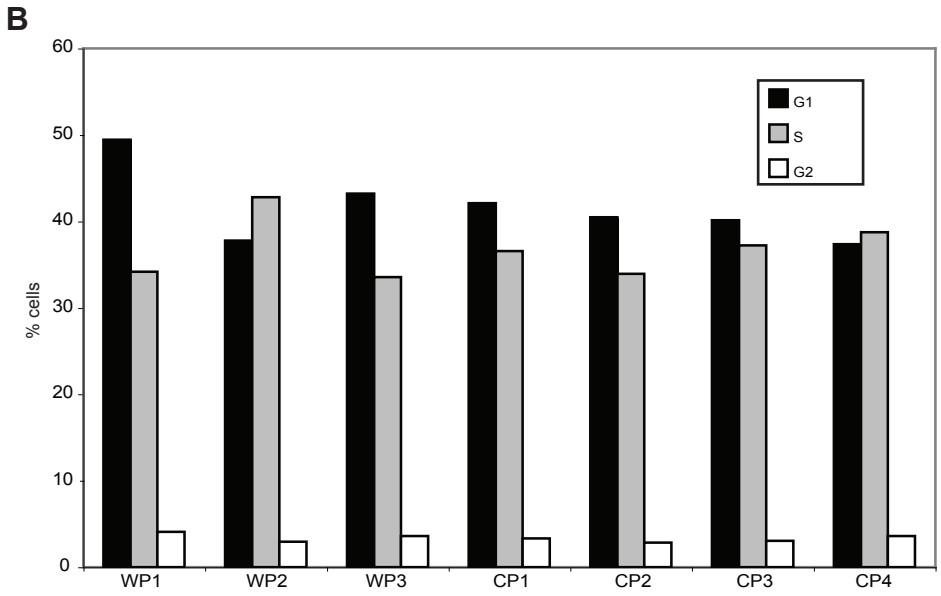


Figure 3.4 – FACS analysis of W1, CN1 and CN2.
A, Histogram of DNA content measured by FACS. Cells were excited at 488nm and the fluorescence measured at 562-588nm, the level of fluorescence giving an estimate of DNA content in PI stained cells. **B**, Bar chart of the percentage of cells in each cell cycle stage, calculated from the DNA content, including p-values from χ^2 tests.





C

Observed \ Expected	WP1	WP2	WP3	CP1	CP2	CP3	CP4
WP1	1	0.0556	0.6157	0.4449	0.2822	0.2516	0.1054
WP2	0.0724	1	0.1873	0.463	0.2851	0.6126	0.7596
WP3	0.6538	0.2311	1	0.8564	0.8187	0.7047	0.4493
CP1	0.4941	0.4868	0.8521	1	0.8395	0.9371	0.6873
CP2	0.3679	0.3606	0.8437	0.8521	1	0.8564	0.5975
CP3	0.3182	0.644	0.7012	0.9418	0.8437	1	0.8395
CP4	0.1645	0.7634	0.4516	0.7082	0.5655	0.8353	1

Figure 3.5 – FACS analysis of WP1, WP2, WP3, CP1, CP2, CP3 and CP4.
A, Histogram of DNA content measured by FACS. Cells were excited at 488nm and the fluorescence measured at 562-588nm, the level of fluorescence giving an estimate of DNA content in PI stained cells. **B**, bar chart of the percentage of cells in each cell cycle stage, calculated from the DNA content. **C**, p-values obtained from χ^2 tests, comparing cell lines pair-wise. Wildtype cell lines are highlighted in black and CdLS cell lines are not highlighted. P-values ≤0.05 that denote a significant difference are highlighted in grey.

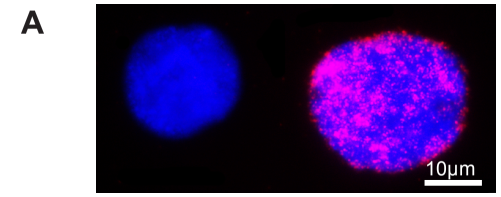
the histogram does not alter (Figure 3.5A). It is from the shape of the histogram, and not the position of the histogram, that FlowJo creates gates for cell cycle stage. CdLS cell lines appear to have taken up proportionately more PI than wildtype; as PI is a dye that intercalates with the DNA, this may in itself indicate that the chromatin structure of CdLS cell lines is altered in CdLS cell lines.

3.3 The increase in nuclear area is not due to cells spending more time in G_2

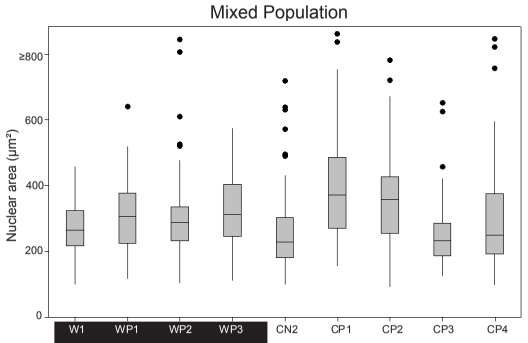
Nuclear area increases as DNA content increases, so a higher percentage of cells in G_2 (Figure 3.4) might have accounted for the increase in nuclear area observed in CN1 and CN2.

To determine if the nuclear area change was due to the cell cycle change in CdLS cells, I decided to study the nuclear area at each cell cycle stage. I carried out a two-hour BrdU incorporation on the four control cell lines (Section 2.6.6), W1 GIA 001S, WP3 and WP2, and five of the CdLS cell lines, CN2, CP1, CP2, CP3 and CP4, then fixed the cells in MAA. The cells were dropped onto slides as with the previous nuclear area analysis, then treated with RNase and denatured before detection of BrdU using a mouse anti-BrdU antibody, followed by a Texas-red conjugated anti-mouse antibody. The slides were then mounted in DAPI and examined using fluorescence microscopy.

Cells which I judged as positive for BrdU staining were considered to have undergone DNA replication at some point within the two hour period of treatment, so I judged these cells to be in S-phase or G_2 (Figure 3.6A). Those that were negative for BrdU staining were judged to be in G_1 as their DNA had not replicated within the time point (Figure 3.6A). Of the few metaphase spreads that I observed, most were negative for BrdU staining, though some had a few loci with BrdU staining, most likely the late replicating loci. Because of this I thought it unlikely that any nuclei with BrdU incorporation had progressed through mitosis into G_1, and similarly unlikely that any G_2 cells had replicated their DNA prior to the BrdU incorporation, so the two-hour incubation was a suitable window. I took 50 images of G_1 nuclei

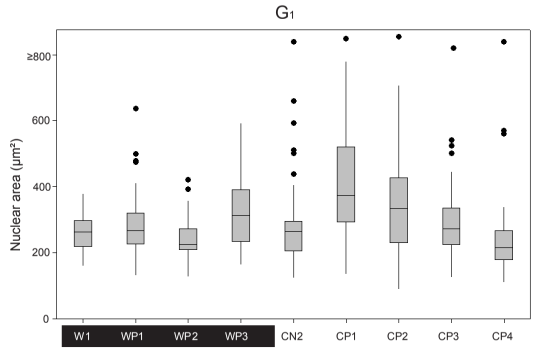


B



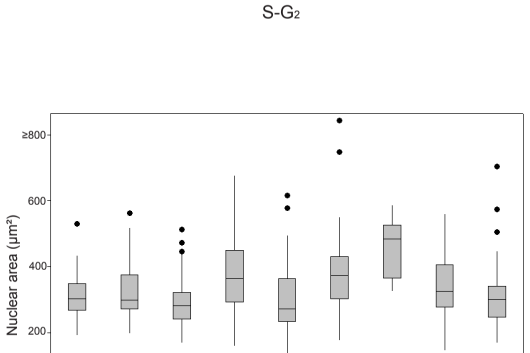
	W1	WP1	WP2	WP3	CN2	CP1	CP2	CP3	CP4
W1	1	0.0136	0.1345	0.0003	0.0298	0	0	0.0047	0.9503
WP1	0.0136	1	0.3487	0.2365	0.0001	0.0001	0.0133	0	0.0642
WP2	0.1345	0.3487	1	0.354	0.0004	0	0.0008	0	0.2163
WP3	0.0003	0.2365	0.354	1	0	0.0048	0.1794	0	0.003
CN2	0.0298	0.0001	0.0004	0	1	0	0	0.8183	0.0845
CP1	0	0.0001	0	0.0048	0	1	0.1511	0	0
CP2	0	0.0133	0.0008	0.1794	0	0.1511	1	0	0.001
CP3	0.0047	0	0	0	0.8183	0	0	1	0.0347
CP4	0.9503	0.0642	0.2163	0.003	0.0845	0	0.001	0.0347	1

C



	W1	WP1	WP2	WP3	CN2	CP1	CP2	CP3	CP4
W1	1	0.4689	0.0736	0.0127	0.9698	0	0.0016	0.2273	0.006
WP1	0.4689	1	0.0156	0.1894	0.5357	0	0.0272	0.6751	0.0012
WP2	0.0736	0.0156	1	0.0007	0.1268	0	0.0001	0.0103	0.1868
WP3	0.0127	0.1894	0.0007	1	0.0946	0.0038	0.4503	0.2496	0.0001
CN2	0.9698	0.5357	0.1268	0.0946	1	0	0.0137	0.3289	0.02
CP1	0	0	0	0.0038	0	1	0.0682	0.0002	0
CP2	0.0016	0.0272	0.0001	0.4503	0.0137	0.0682	1	0.0565	0.0001
CP3	0.2273	0.6751	0.0103	0.2496	0.3289	0.0002	0.0565	1	0.0015
CP4	0.006	0.0012	0.1868	0.0001	0.02	0	0.0001	0.0015	1

D



	W1	WP1	WP2	WP3	CN2	CP1	CP2	CP3	CP4
W1	1	0.5432	0.1467	0.0005	0.1606	0.0004	0.0004	0.0688	0.6075
WP1	0.5432	1	0.0914	0.0101	0.0839	0.0055	0.0022	0.2756	0.3131
WP2	0.1467	0.0914	1	0.0001	0.7433	0.0001	0.003	0.0082	0.5147
WP3	0.0005	0.0101	0.0001	1	0.0003	0.9533	0.0701	0.0693	0.0005
CN2	0.1606	0.0839	0.7433	0.0003	1	0.0003	0.0018	0.0193	0.5556
CP1	0.0004	0.0055	0.0001	0.9533	0.0003	1	0.0595	0.0946	0.0006
CP2	0.0004	0.0022	0.003	0.0701	0.0018	0.0595	1	0.0083	0.0011
CP3	0.0688	0.2756	0.0082	0.0693	0.0193	0.0946	0.0083	1	0.0594
CP4	0.6075	0.3131	0.5147	0.0005	0.5556	0.0006	0.0011	0.0594	1

Figure 3.6 – BrdU staining demonstrates that the increase in nuclear area is unrelated to the cell cycle.

A, BrdU stained cells. Detection of BrdU after a two-hour BrdU pulse; BrdU incorporation was detected in Texas Red and the nuclei counterstained with DAPI. The cell on the left is negative for BrdU staining so was scored as being in G1, and the cell on the right is positive for BrdU staining so was scored as being in G2 or S-phase. **B**, Boxplot showing the distribution of nuclear area measurements in asynchronous cells (n=100), which were not scored for BrdU staining, as **Figure 3.2**. The table shows p-values obtained from Mann-Whitney U-tests, comparing cell lines pair-wise. Wildtype cell lines are highlighted in black and CdLS cell lines are not highlighted. P-values ≤0.05 that denote a significant difference are highlighted in grey. **C**, as **B**, but for cells scored as being in G1 (BrdU-; n=50). **D**, as **B**, but for cells scored as being in S-G2 (BrdU+; n=50).

and 50 S-G_2 images for further study. I analysed the nuclear area as before, (Section 3.1), to compare the nuclear area of G_1 cells to S-G_2 cells. I also examined 100 images of each cell line without looking for BrdU staining, so that I could compare each cell cycle stage to a mixed cell stage population (asynchronous), that would reflect all cell cycle stages in their normal proportions, and would also allow me to repeat analysis of nuclear size between the old and new cell lines. Within each cell cycle stage, I compared all the cell lines pair-wise using Mann-Whitney U-tests.

The asynchronous population did not follow the same pattern as the previous data on nuclear area (Figure 3.6B; Figure 3.2; Figure 3.3). Mostly, the wildtype cell lines did not have significantly different nuclear areas to one another. However the CdLS cell lines varied in nuclear area, with some significantly larger than wildtype (CP1 and CP2) and some with a similar size to, or even smaller than wildtype (CN2, CP3 and CP4). It is unclear why these results differ from those previously gathered (Figure 3.2; Figure 3.3).

The G_1 population follows a similar pattern to the asynchronous population (Figure 3.6C). There is more variation within the wildtype cell lines than in the mixed population. In the G_1 population CN2 has similar nuclear area to all wildtype cell lines and CP4 has statistically smaller nuclei than all cell lines, except WP2. CP1, CP2 and CP3 have large nuclei, which are mostly statistically different to wildtype.

The S-G_2 population also shows a similar pattern to the asynchronous and G_1 populations (Figure 3.6D). Three of the wildtype cell lines, W1, WP1 and WP2, plus two CdLS lines, CN2 and CP4 do not have significantly different nuclear areas. CP1, CP2 and CP3 have significantly larger nuclei than all members of the other group, but of these three, only CP2 and CP3 are significantly different to each other.

In this experiment CN2, CP3 and CP4 behaved more like wildtype cell lines than CdLS, however CP1 and CP2 had consistently larger nuclei than wildtype, in the mixed population and in both cell cycle stages. Although the data in this experiment were not consistent with data from previous experiments, the data are consistent between the cell cycle stages in this experiment, so we can consider that any change in nuclear area in CdLS is independent of the change in the cell cycle of CdLS.

Despite this, I conclude that gross nuclear area, as assessed cytologically, is too irreproducible to explore further.

3.4 Analysis of nuclear size by FACS

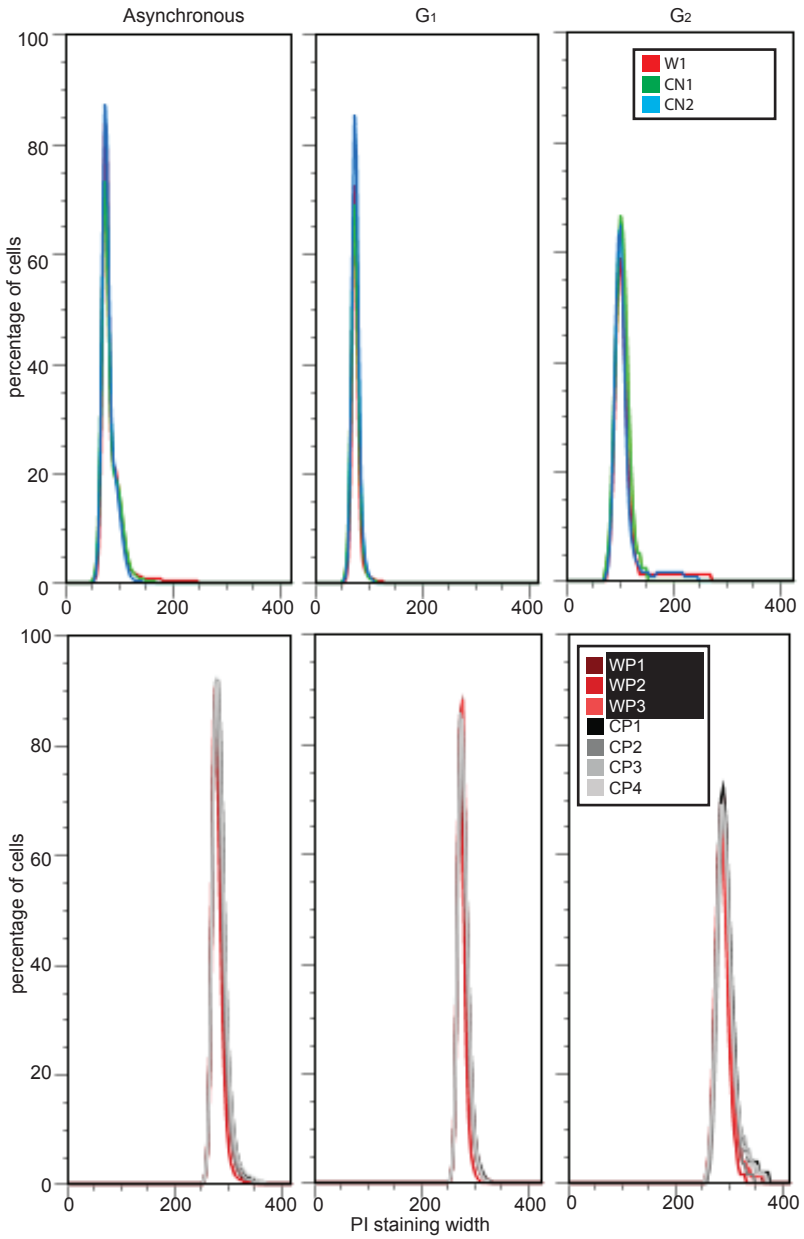
As I was not able to reproduce the change in nuclear size in CdLS cell lines in the BrdU experiment, I wanted to assess nuclear size by a different method. Nuclear size can also be estimated from FACS data, by measuring the width of the PI signal, with larger nuclei giving a higher relative width. This method is able to identify the size of nuclei, rather than the size of cells, as it relies upon the PI signal, which intercalates with the DNA (Sharpless and Melamed, 1976; Kang *et al*, 2010).

I measured the fluorescence width from my previous FACS data, looking at asynchronous, G_1 and G_2 populations, in all the CdLS and wildtype cell lines previously studied. I found that in neither the Newcastle nor the Philadelphia cell lines, was there a difference in PI signal width between the CdLS and wildtype cell lines, and that this was consistent through the cell cycle (Figure 3.7). Due to the much higher sample sizes used in FACS analysis, and the lack of human involvement in choosing cells to analyse, I believe that that FACS PI signal width gives a much more accurate measure of nuclear size than microscopic analysis. Based on this assumption, I suggest that there is no change in nuclear size in CdLS cells compared to wildtype.

3.5 Discussion

The CdLS cell lines I have examined do not show such a consistent increase in nuclear area as was seen when the condensin subunits SMC2 and SMC4 were knocked down in mouse ES cells. However SMC1A was also tested in this RNAi screen, and a change in nuclear size was not detected (Fazzio and Panning, 2010). It is important to consider the protein levels in the two experiments, knockdown of *Smc2* in the published experiment is almost complete, whereas CdLS cells show only 30% loss of NIPBL protein (Liu *et al*, 2009).

A



B

	Mean PI width		
Cell line	Asynchronous	G_1	G_2
W1	88.07	76.53	119.76
CN1	83.84	72.98	106.95
CN2	79.73	70.79	112.39
WP1	279.28	274.76	289.35
WP2	276.49	271.52	284.24
WP3	280.33	275.86	289.32
CP1	282.05	275.94	294.22
CP2	279.53	275.13	291.1
CP3	280.4	274.99	292.8
CP4	282.48	275.8	292.04

Figure 3.7 - Nuclear size as shown by PI staining width.
Histograms of PI staining width, measured by FACS. Cells were excited at 488nm and the fluorescence measured at 562-588nm, the width of fluorescence giving an estimate of nuclear size in PI stained cells. **A,** Fluorescence width in seven CdLS LCLs and four wildtype LCLs, in asynchronous, G_1 and G_2 cells. **B**, Mean PI width of the seven CdLS LCLs and four wildtype LCLs

Because I could detect no overall gross change in chromatin compaction, as assessed by nuclear size, I next analysed chromatin compaction at specific loci.

4 Chromatin condensation at specific regions in CdLS

In the previous chapter, analyses using the cell lines W1, CN1 and CN2, had suggested that mutation of *NIPBL* in CdLS may cause a general decompaction of chromatin. To determine if chromatin compaction was altered at specific genomic regions with differing characteristics in these cells, I then set out to examine defined genomic locations via study of chromosome territories and extensive 2D FISH analysis. I also examined some of these regions at different stages of the cell cycle, and via 3D FISH.

4.1 Chromosome territory size in CdLS

I was interested in the sizes of the CTs in CdLS as it has been shown that changes in compaction of the chromatin result in an alteration of the size of the CTs. For example, knockdown of BAF53, which is a component of a number of chromatin remodelling complexes, including SWI/SNF and TIP60, causes decompaction of chromatin and enlargement of CTs (Lee *et al*, 2007).

I measured CT area by hybridising a DIG-11-dUTP-labelled chromosome paint to MAA fixed nuclei and detecting the CTs in FITC, with the nuclei in DAPI. I was then able to measure the size of the CTs and nuclear area using an IPLab script (Section 2.11.2.2; Figure 2.1). I normalised the CT area to the nuclear area, then compared wildtype to CdLS cells using Mann Whitney U-tests, and considered data to be significantly different where $p \leq 0.05$.

I measured the CT size of four chromosomes, 11, 13, 17 and 18; chromosome 11 area was measured in W1 (wt) cells compared to CN1 (CdLS) and CN2 (CdLS) and the areas of chromosomes 13, 17 and 18 were compared between W1 and CN1 cells. I was interested in chromosomes 13 and 18 because both are gene-poor chromosomes that are located close to the nuclear periphery and have compact

chromatin (Gilbert *et al*, 2004). In contrast, chromosome 17 is gene-rich, has open chromatin and located towards the nuclear centre (Boyle *et al*, 2001). Chromosome 11 was of interest because it locates towards the nuclear periphery, although it is gene-rich (Boyle *et al*, 2001). I was also carrying out other studies on looping out from chromosome 11 (Section 4.2), and looking at compaction at a specific region on chromosome 11, 11q13, by 2D FISH (Section 4.3.1). It is important to carry out these studies on a number of different chromosomes, with varied characteristics, as CT size varies, not just with chromosome size, but also with compaction of the CT due to histone acetylation and other factors (Croft *et al*, 1999).

I looked at the CTCF and cohesin binding density of each of the chromosomes studied (Figure 4.1A; Figure 4.1B; Barski *et al*, 2007; Liu *et al*, 2009). I found that chromosome 11 had a higher than average CTCF binding density but lower than average cohesin binding density, chromosome 17 had higher than average CTCF and cohesin binding densities and chromosomes 13 and 18 had lower than average CTCF and cohesin binding (Figure 4.1B).

I found that the CT size of chromosome 11 was significantly increased in both CN1 and CN2 compared to W1 ($p<0.02$; Figure 4.2C). The CT area of chromosome 13 did not significantly alter in CN1 compared to W1 ($p=0.8288$; Figure 4.2C). The CT areas of chromosomes 17 and 18 both significantly decreased in CN1 compared to W1 ($p=0.001$ and $p=0.035$ respectively; Figure 4.2C). There appeared to be no correlation between the binding density of CTCF and cohesin, and the change in CT size in CdLS.

I then looked at the physical size of the CTs compared to the genomic length of the chromosomes (Figure 4.2B). I considered that there would be a correlation between the physical size of the chromosome and the genomic length, but that many chromosomes would deviate strongly from this relationship based upon gene density and chromatin topographical features. I therefore compared the mean CT sizes normalised to nuclear area that I measured, to the genomic size of the chromosome normalised to the size of the whole genome (Croft *et al*, 1999). I found that there was some correlation between chromosome size and CT size in W1 cells, but a much stronger correlation in CN1 cells (Figure 4.2B). This suggests that the effect of

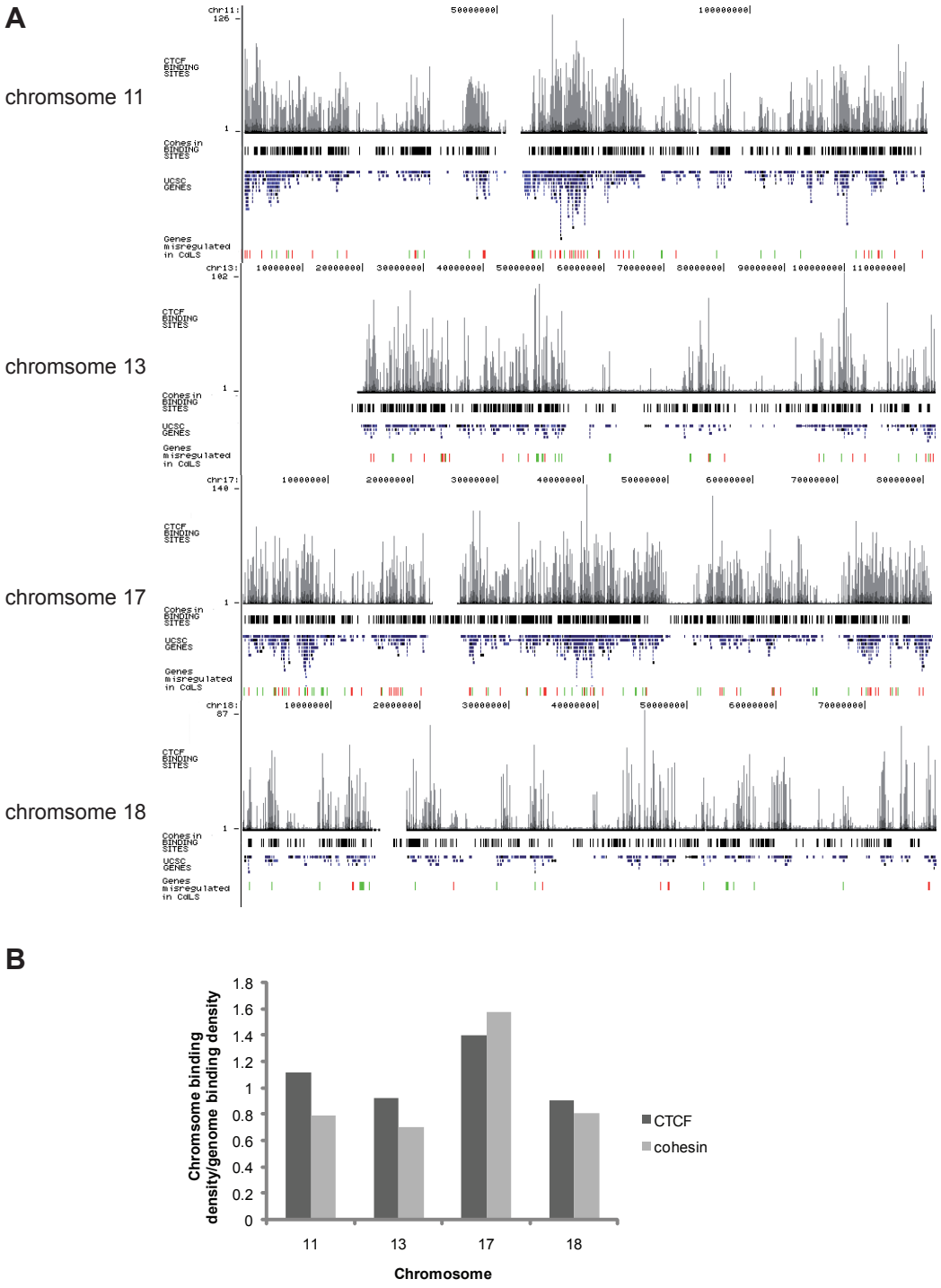


Figure 4.1 - CTCF and cohesin binding on human chromosomes 11, 13, 17 and 18. **A,** UCSC tracks of the four chromosomes studied (GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). The position in bp is indicated at the top. CTCF binding across the chromosome is indicated as a bar chart in black (Barski *et al*, 2007). Cohesin binding across the chromosome is indicated below in black (Liu *et al*, 2009). The genes are in blue, and genes that are misregulated in CdLS LCLs are highlighted in red and green below (Liu *et al*, 2009). **B**, Bar-chart showing the CTCF and Rad21 binding density of each chromosome (Barski *et al*, 2007; Liu *et al*, 2009).

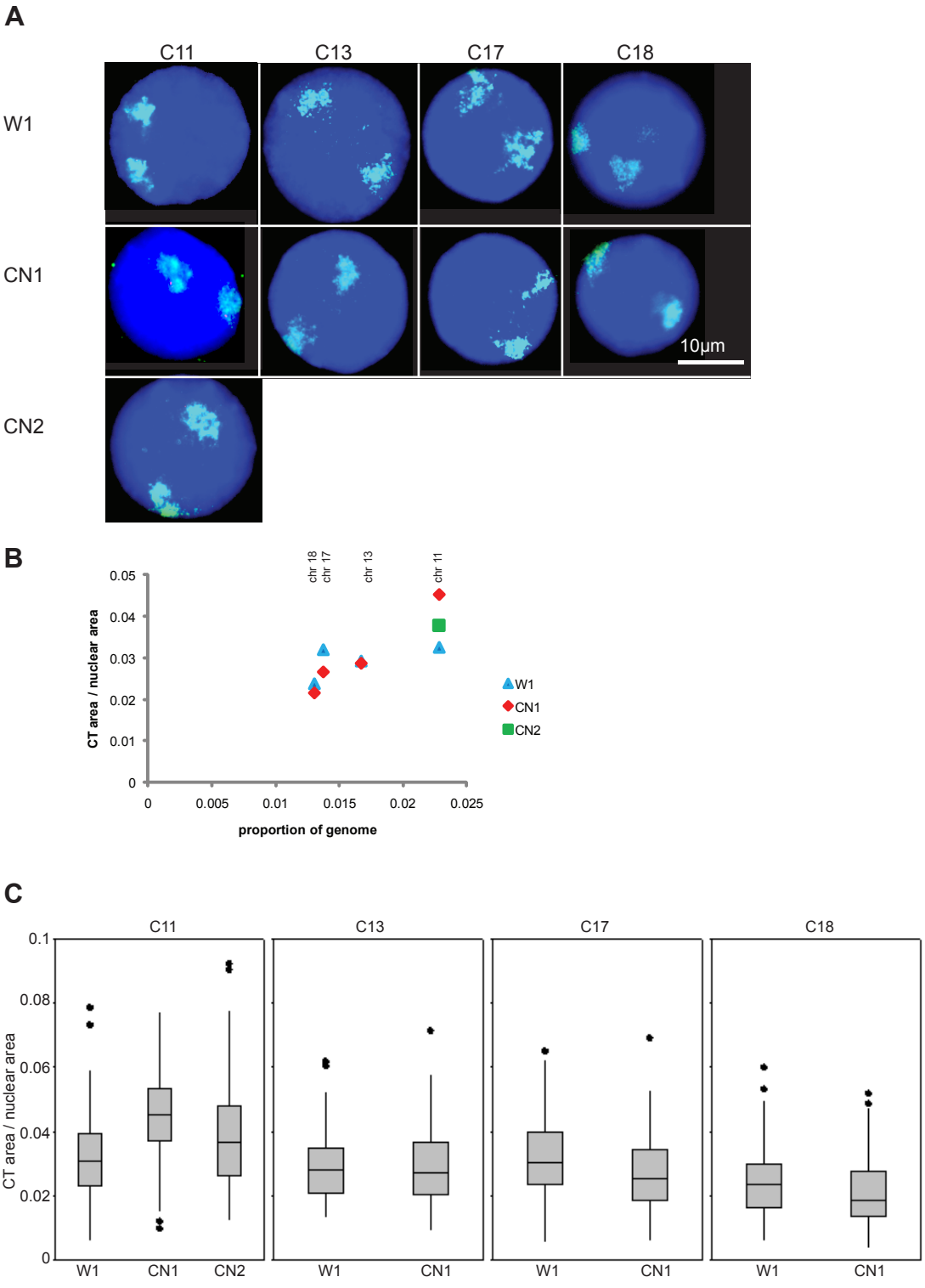


Figure 4.2 - Chromosome territory size in W1, CN1 and CN2.
A, sample fluorescence microscopy images of chromosome territories in nuclei. DNA is stained with DAPI and chromosome territories are stained in green. **B**, Scatter-plot showing the relationship between the normalised CT area in W1, CN1 and CN2 and the proportion of the genome that that chromosome makes up, for chromosomes 11, 13, 17 and 18. **C**, Boxplots showing the territory size of chromosome 11 in W1, CN1 and CN2 , and chromosomes 13, 17 and 18 in W1 and CN1, normalised to nuclear area, as **Figure 3.2**, n=100

CdLS is to homogenise the chromatin such that on a large scale, all chromatin will tend towards having a linear relationship between genomic length and space taken up.

4.2 Intra-chromosome territory organisation in CdLS

I then measured the intra-chromosome territory organisation of two regions of chromosome 11 from the CT. It has been shown on various chromosomes that genes loop out of their CT upon activation (Mahy *et al*, 2002b; Chambeyron and Bickmore, 2004; Morey *et al*, 2007). It is thought that this reflects their association with transcription factories, forming ACHs (Kurukuti *et al*, 2006; Chernukhin *et al*, 2007; Mitchell and Fraser, 2008; Heard and Bickmore, 2007); it is possible that the formation of these ACHs may be mediated by cohesin. It is important to note that looping out is not sufficient to activate transcription (Morey *et al*, 2009).

Looping out from CTs was measured by hybridising a DIG-11-dUTP-labelled paint to the CT (as Section 4.1) and hybridising a biotin-16-UTP-labelled probe to a specific region, then detecting the CT in FITC and the probe in Texas Red. The shortest distance to a chromosome territory edge was then measured using an IPLab script, giving negative values for probes outside the CT, and positive values for probes inside the CT. The distance between the probe and CT edge was normalised to the nuclear radius, estimated from the measured nuclear area (Section 2.11.2.3; Figure 2.2). I then compared CdLS cell line datasets to wildtype using Mann-Whitney U-tests, and considered datasets to be significantly different where $p \leq 0.05$.

I measured the intra-territory organisation of two regions of chromosome 11, 11p13 and 11q13, measured in W1, CN1 and CN2 at 11q13 and in W1 and CN1 only at 11p13 (Figure 4.3A). 11p13 has been shown to be gene-poor and normally located well within the CT (Mahy *et al* 2002a; Mahy *et al*, 2002b). Previous work in the lab had demonstrated that 11q13 is a very open chromatin region that is frequently located outside the chromosome territory (Gilbert *et al*, 2004; Mahy *et al*, 2002b). This region is both gene dense and very transcriptionally active, and is considered to be a region of increased gene expression (ridge) of irregularly-shaped, decondensed chromatin located towards the nuclear centre (Goetze *et al*, 2007;

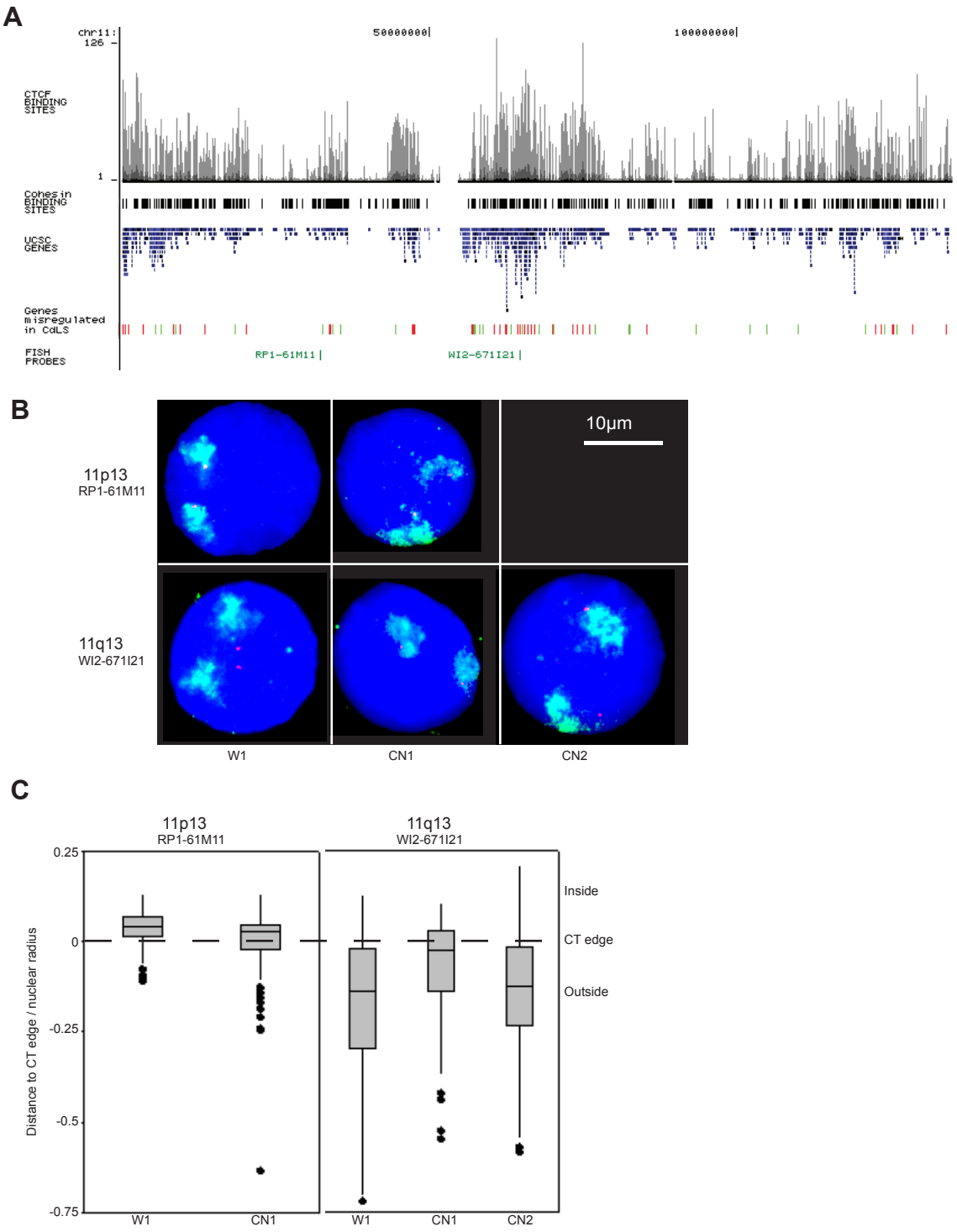


Figure 4.3 - Looping out of chromosome 11 in W1, CN1 and CN2.
A, UCSC track of human chromosome 11 (GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). The position in bp is indicated at the top. CTCF binding across the chromosome is indicated as a bar chart in black at the (Barski *et al*, 2007). Cohesin binding across the chromosome is indicated below in black (Liu *et al*, 2009). The genes are in blue, and genes that are misregulated in CdLS LCLs are highlighted in red and green below (Liu *et al*, 2009). The two probes used are indicated below in green, RP1-61M11 at 11p13 and WI2-671I21. **B**, sample fluorescence microscopy images, the DNA is labelled with DAPI, the CTs in green and the probes in red **C**, Boxplots showing distance from the FISH probes to the chromosome territory edge normalised to the nuclear radius, negative values indicate that the probe is outside the chromosome territory and positive values indicate that it is inside, as **Figure 3.2**, n=100.

Gierman *et al*, 2007). More recently, looping out of this region was shown to be dynamic (Müller *et al*, 2010). I used a cosmid, RP1-61M11, to study looping out at 11p13 and the fosmid probe, WI2-671I21, to study looping at 11q13 (Figure 4.3A).

RP1-61M11 at 11p13 is mostly found inside the CT in both W1 and CN1, but is found closer to the territory edge, and even outside the CT in CN1 ($p<0.002$). I found that WI2-671I21 at 11q13, which is mostly found outside the CT in all three cell lines, is located closer to the CT edge in CN1 compared to W1 ($p<0.001$) but that there is no difference in the localisation of WI2-671I21 compared to the CT between W1 and CN2 ($p=0.342$; Figure 4.3B). These data also suggest that CdLS homogenises the chromatin organisation, and that the organisation of chromatin within a CT is tending towards a more random state, as control of large scale looping is lost.

4.3 2D FISH to study CdLS at specific chromatin regions

2D FISH can be used to give a measure of chromatin compaction over a specific region of the genome. It is a well-established technique in the Bickmore lab, as well as others, and has been used to demonstrate changes and variation in compaction. It has been used to compare compaction between different genomic regions in the same cells (Yokota *et al*, 1997; Gilbert *et al*, 2004), to compare the same genomic region through cell differentiation (Chambeyron *et al*, 2004) and to compare the same region in wildtype versus mutant cells (Eskeland *et al*, 2010).

2D FISH has been shown to be a robust method for determining compaction of chromatin that accurately reflects the 3D structure of chromatin, though compared to 3D FISH, 2D FISH generally gives greater distances (Morey *et* al, 2007; Eskeland *et al*, 2010). It is more suitable for examining large datasets than 3D FISH, as imaging in 2D is faster and algorithms for measuring distances are well established. Therefore I only used 3D FISH to confirm my 2D FISH data, and not to carry out initial analysis.

The physical distance between genomic regions of known genomic separation (kb) was measured using pairs of fosmid clones that correspond to a region of interest and that are of known distance apart. The probes are selected by examining the region of interest on the UCSC browser, and choosing fosmid probes from the *fosmid end pairs* track on the March 2006 human genome assembly (Kent et al, 2002; http://genome.ucsc.edu/). One of the clones is labelled with DIG-11-dUTP and the other with biotin-16-dUTP, then hybridised to MAA fixed cells on slides. I then detected DIG-11-dUTP with fluorescein-conjugated antibodies and biotin-16-dUTP with Texas Red-conjugated antibodies (Section 2.8; Figure 4.4B, 4.6B, 4.8B and 4.10B). I measured the physical distance between the two probes (Figure 2.3) and determined the mean squared separation (r^2), as there is a linear relationship between this and the genomic distance (Figure 4.4). The r^2 gives an estimate of the chromatin compaction, as larger values at the same genomic distance suggests that the chromatin is less compact. It is possible to use 2D FISH to measure chromatin compaction at genomic distances of <1Mb, as chromatin appears to act as a polymer that follows a random-walk giant-loop model, with loops above 1.5Mb (Sachs *et al*, 1995; van den Engh *et al*, 1992). The data is also normalised to the nuclear area (r^2/a), as nuclear size varies based not only on the cell type, as seen in the previous chapter, but on the humidity on the day of preparation. The distribution of r values should follow a Rayleigh distribution, where the data have a discrete cut-off at zero to the left of the distribution, but extend towards infinity to the right of the distribution. As a result the ratio of median to mean is ~0.9 and the ratio of standard deviation to mean is ~0.5 (Figure 1.6; Figure 4.4). For a more linear confirmation the median to mean ratio is approximately 1, and the ratio of standard deviation to mean is around 0.25.

I first studied chromatin compaction by FISH in the three original cell lines, W1, CN1 and CN2, and at four regions of the genome. Analysis of Variance (ANOVA) was carried on $\log_{10}$ of r^2 or r^2/a on each single experiment, comparing datasets from W1, CN1 and CN2 under each condition to give an indication of whether there was variation between the three datasets. Variation was considered significant when this gave a p value ≤ 0.05. This analysis also generated pairwise Tukey's confidence intervals, and pairs of datasets were considered to be significantly different when

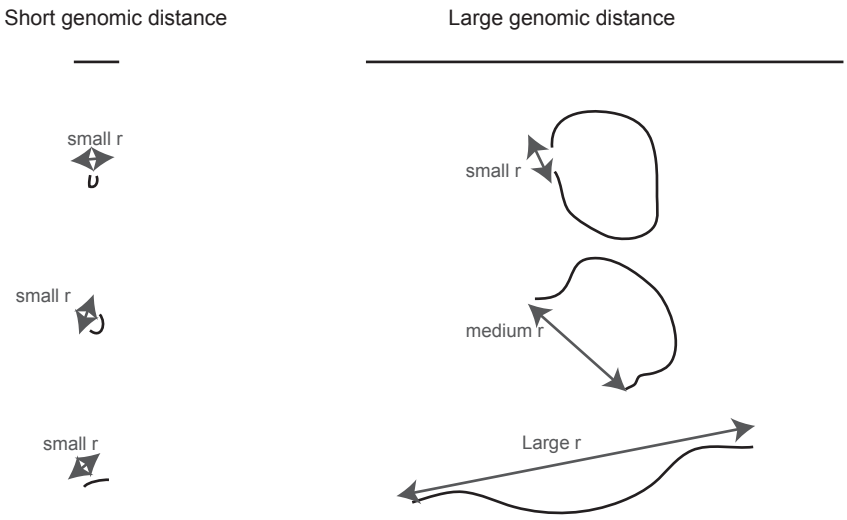


Figure 4.4 - Mean r² is proportional to genomic distance.
An illustration showing how the mean r² measured in a FISH experiment is proportional to the genomic distance. If we compared mean r to genomic distance, we would find that over large genomic separations the small values of r that occur by chance because of the random folding fibre would significantly reduce the mean r. However if we compare the mean r² we find that the larger r² values in the dataset are sufficiently high to increase the mean. At shorter genomic separations, r is small and varies very little, and squaring it makes little difference. As a result r² is proportional to genomic distance.

these confidence intervals did not overlap one another. I then carried out Mann Whitney U-tests comparing the datasets from the two CdLS cell lines to the W1 dataset, and considered data to be significantly different where $p \leq 0.05$.

4.3.1 Chromatin decompaction at regions of high CTCF binding density in CdLS cell lines

I first selected a candidate region at 11q13 for studying chromatin compaction, where I had studied looping-out of the CT (Section 4.2). This is a gene-rich ridge region, located outside the chromosome territory and towards the nuclear centre (Figure 4.5A; Gilbert *et al*, 2004; Mahy *et al*, 2002b; Goetze *et al*, 2007; Gierman *et al*, 2007). Since cohesin and CTCF co-localise (Wendt *et al*, 2008; Parelho *et al*, 2008; Stedman *et al*, 2008), I examined CTCF binding at 11q13 from published data (Barski *et al*, 2007) and found that it binds very densely, at approximately 2798 sites/Mb, compared to the global average binding of 906 sites/Mb estimated from the same dataset. Also, after selecting the region for study, it was subsequently revealed that this region also contains a small number of genes that are misregulated in CdLS LCLs and that there are 11 cohesin binding sites across the 1Mb region studied in LCLs as identified by ChIP-chip for Rad21 ($p=1 \times 10^{-6}$ compared to an input sample; Liu *et* al, 2009; Figure 4.5A).

Probe pairs were selected at separations of ~250kb, ~500kb, ~750kb and ~1Mb; WI2-671I21 was selected as an anchor probe at the centromere proximal end of the region and labelled with DIG-11-dUTP whilst WI2-1737E8, WI2-412A22, WI2-2372A13 and WI2-3325E18 were 249kb, 517kb, 750kb and 996kb away respectively and labelled with biotin-16-dUTP (Table 2.7; Figure 4.5A; Figure 4.5B). 1Mb was chosen as the maximum distance as this would represent a separation where the random-walk model would hold (van den Engh *et al*, 1992) and ~250kb was selected as a minimum distance as below this distance probes may have been hard to resolve by light microscopy.

There is an increase in interphase separation observed with the increase in genomic distance at the 11q13 region. The relationship between r^2 and genomic distance is close to linear up to 750kb, but the linearity breaks down at the 1Mb

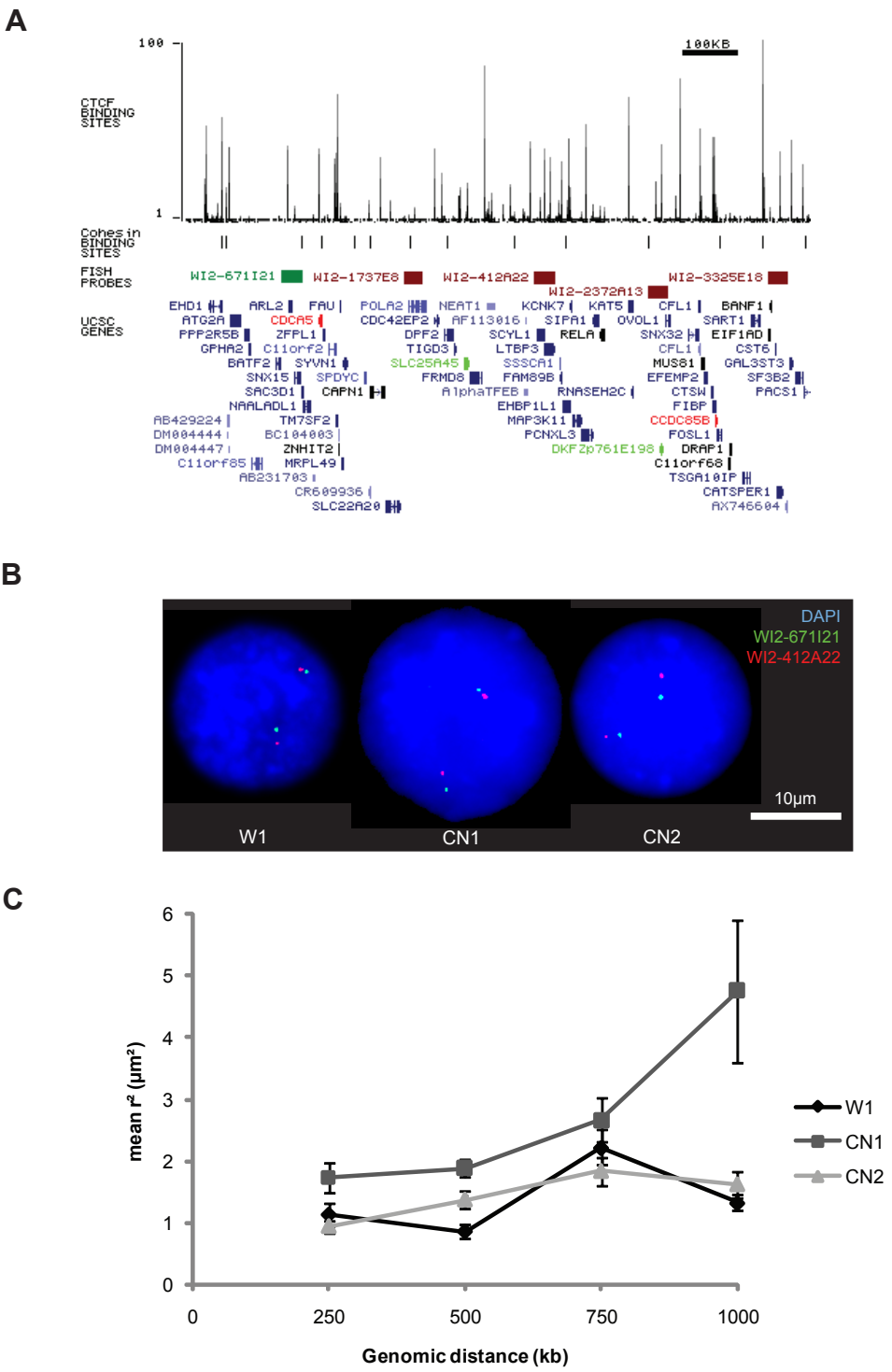


Figure 4.5 – Chromatin is decompacted at 11q13 in CdLS.
A, UCSC track of 11q13 (chr11:64,500,000-65,900,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). CTCF binding across the region is indicated as a bar chart in black at the top of the track (Barski *et al*, 2007). Cohesin binding across the region is indicated below in black (Liu *et al*, 2009). The positions of the FISH probes used are indicated below, the reference probe is indicated in green and the different distance probes are indicated in red. The genes in the region are in blue, however those that are upregulated in CdLS LCLs compared to wildtype are highlighted in green, and those downregulated in red (Liu *et al*, 2009). **B**, sample fluorescence microscopy images of nuclei hybridised with probes separated by ~500kb. **C**,The mean interphase separation squared is plotted against the genomic distance as measured in W1, CN1 and CN2 cell lines. The error bars represent the standard error of the mean (SEM).

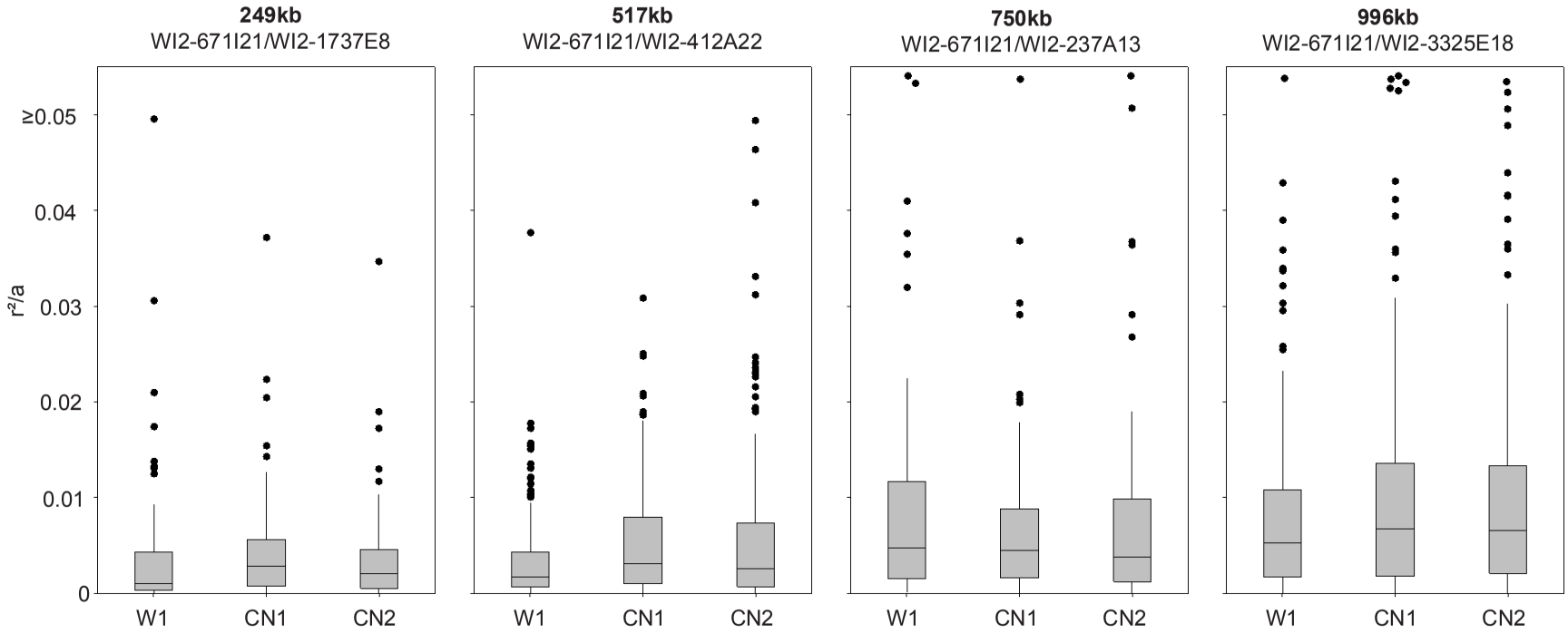


Figure 4.6 – Chromatin is decompacted at 11q13 in CdLS.
Boxplots showing the normalised interphase separation squared (r^2/a) of 11q13, at the four genomic distances in the three cell lines, as **Figure 3.2**, n=100

Table 4.1 - 2D FISH r^2 statistics

The statistics of every r^2 dataset studied by 2D FISH. Significant differences are highlighted in grey.

Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2 (μm^2)	Median r^2 (μm^2)	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?	median r/ mean r	stdev r/ mean r
11q13	WI2-671I21	WI2-1737E8	249	W1	1.14	0.41	0.001	-		0.77	0.80
				CN1	1.74	1.04		0.001	no	0.92	0.65
				CN2	0.94	0.49		0.697	yes	0.86	0.67
	WI2-671I21	WI2-412A22	517	W1	0.87	0.35	0.000	-		0.81	0.79
				CN1	1.89	1.00		0.000	no	0.86	0.64
				CN2	1.37	0.59		0.000	yes	0.81	0.69
	WI2-671I21	WI2-2372A13	750	W1	2.23	1.18	0.115	-		0.86	0.64
				CN1	2.67	1.41		0.304	no	0.86	0.63
				CN2	1.84	1.13		0.330	yes	0.93	0.62
	WI2-671I21	WI2-3325E18	996	W1	1.33	0.74	0.002	-		0.87	0.60
				CN1	4.76	1.61		0.000	no	0.79	0.92
				CN2	1.62	0.75		0.711	yes	0.82	0.68
EDC	WI2-2080I19	WI2-2423D6	262	W1	0.53	0.16	0.000	-		0.74	0.89
				CN1	0.92	0.33		0.000	no	0.81	0.90
				CN2	0.64	0.27		0.003	no	0.81	0.73
	WI2-2080I19	WI2-553L9	552	W1	1.24	0.88	0.007	-		0.95	0.52
				CN1	1.98	1.29		0.002	no	0.91	0.52
				CN2	1.79	1.08		0.057	yes	0.88	0.55
	WI2-2080I19	WI23057F14	741	W1	1.64	0.82	0.001	-		0.86	0.69
				CN1	2.64	1.59		0.002	no	0.91	0.63
				CN2	2.44	1.55		0.005	no	0.93	0.60
	WI2-2080I19	WI2-784N16	1055	W1	1.95	1.11	0.018	-		0.89	0.62
				CN1	3.54	1.65		0.000	no	0.83	0.69
				CN2	2.65	1.23		0.132	yes	0.83	0.69

HoxD	WI2-3163O19	WI2-1038D9	253	W1	1.02	0.09	0.000	-		0.55	1.54
				CN1	0.18	0.04		0.000	no	0.65	1.00
				CN2	0.45	0.18		0.000	no	0.81	0.74
	WI2-3163O19	WI2-2744M3	396	W1	0.61	0.29	0.168	-		0.84	0.71
				CN1	0.71	0.31		0.634	yes	0.81	0.72
				CN2	0.86	0.48		0.021	yes	0.91	0.67
	WI2-3163O19	WI2-1919C15	491	W1	0.74	0.42	0.302	-		0.90	0.66
				CN1	0.84	0.42		0.917	yes	0.82	0.75
				CN2	0.88	0.50		0.290	yes	0.89	0.64
	WI2-3163O19	WI2-1957H23	742	W1	0.80	0.32	0.001	-		0.78	0.72
				CN1	1.17	0.74		0.001	no	0.93	0.60
				CN2	0.61	0.36		0.681	yes	0.90	0.62
	WI2-3163O19	WI2-1241H8	1007	W1	0.79	0.34	0.243	-		0.80	0.72
				CN1	0.86	0.48		0.117	yes	0.88	0.62
				CN2	1.03	0.41		0.151	yes	0.77	0.69
18q22	WI2-1702P7	WI2-502C21	239	W1	0.46	0.15	0.386	-		0.74	0.82
				CN1	0.40	0.25		0.118	yes	0.92	0.61
				CN2	0.49	0.16		0.495	yes	0.71	0.76
	WI2-1702P7	WI2-1735D16	376	W1	0.55	0.27	0.798	-		0.87	0.74
				CN1	0.60	0.26		0.520	yes	0.81	0.73
				CN2	0.51	0.25		0.925	no	0.87	0.73
	WI2-1702P7	WI2-1708L7	486	W1	2.51	0.38	0.003	-		0.72	1.33
				CN1	0.91	0.24		0.004	no	0.75	0.99
				CN2	0.64	0.26		0.008	no	0.82	0.77

Table 4.2 - 2D FISH r^2/a statistics

The statistics of every r^2/a dataset studied by 2D FISH. Significant differences are highlighted in grey.

Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2/a	Median r^2/a	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?
11q13	WI2-671I21	WI2-1737E8	249	W1	0.0036	0.0011	0.029		
				CN1	0.0043	0.0029		0.007	no
				CN2	0.0035	0.0021		0.178	yes
	WI2-671I21	WI2-412A22	517	W1	0.0034	0.0017	0.020		
				CN1	0.0053	0.0031		0.004	no
				CN2	0.0055	0.0026		0.036	yes
	WI2-671I21	WI2-2372A13	750	W1	0.0091	0.0048	0.444		
				CN1	0.0074	0.0045		0.645	yes
				CN2	0.0066	0.0038		0.224	yes
	WI2-671I21	WI2-3325E18	996	W1	0.0082	0.0053	0.499		
				CN1	0.0160	0.0068		0.261	yes
				CN2	0.0121	0.0066		0.139	yes
EDC	WI2-2080I19	WI2-2423D6	262	W1	0.0025	0.0009	0.002		
				CN1	0.0041	0.0014		0.003	no
				CN2	0.0024	0.0011		0.053	yes
	WI2-2080I19	WI2-553L9	552	W1	0.0050	0.0030	0.113		
				CN1	0.0062	0.0040		0.102	yes
				CN2	0.0052	0.0031		0.907	yes
	WI2-2080I19	WI23057F14	741	W1	0.0057	0.0029	0.002		
				CN1	0.0077	0.0043		0.012	no
				CN2	0.0085	52W19778810664		0.002	no
	WI2-2080I19	WI2-784N16	1055	W1	0.0088	0.0060	0.133		
				CN1	0.0117	0.0068		0.177	yes
				CN2	0.0124	0.0074		0.027	yes

HoxD	WI2-3163O19	WI2-1038D9	253	W1	0.0061	0.0004	0.000		
				CN1	0.0007	0.0002		0.000	no
				CN2	0.0020	0.0006		0.003	yes
	WI2-3163O19	WI2-2744M3	396	W1	0.0026	0.0012	0.056		
				CN1	0.0021	0.0011		0.349	yes
				CN2	0.0044	0.0018		0.088	yes
	WI2-3163O19	WI2-1919C15	491	W1	0.0047	0.0018	0.084		
				CN1	0.0031	0.0016		0.537	yes
				CN2	0.0037	0.0018		0.166	yes
	WI2-3163O19	WI2-1957H23	742	W1	0.0031	0.0014	0.065		
				CN1	0.0035	0.0021		0.118	yes
				CN2	0.0022	0.0012		0.200	yes
	WI2-3163O19	WI2-1241H8	1007	W1	0.0025	0.0012	0.059		
				CN1	0.0026	0.0015		0.316	yes
				CN2	0.0039	0.0017		0.030	no
18q22	WI2-1702P7	WI2-502C21	239	W1	0.0020	0.0007	0.775		
				CN1	0.0013	0.0007		0.688	yes
				CN2	0.0020	0.0007		0.787	yes
	WI2-1702P7	WI2-1735D16	376	W1	0.0021	0.0010	0.316		
				CN1	0.0019	0.0009		0.846	yes
				CN2	0.0025	0.0011		0.286	yes
	WI2-1702P7	WI2-1708L7	486	W1	0.0083	0.0013	0.004		
				CN1	0.0038	0.0009		0.003	no
				CN2	0.0029	0.0011		0.005	no

separation. However in the CdLS cell line CN1 there is still a linear relationship up to 1Mb (Figure 4.5C). This may suggest that a specific cohesin mediated chromatin interaction takes place in between WI2-2372A13 and WI2-3325E18, which is disrupted in CN1 (Figure 4.5A). Despite this none of the datasets significantly deviate from a Rayleigh distribution, suggesting that in all cases the chromatin conforms to a random walk path model (Table 4.1).

At the 11q13 region there is evidence that CN1 has significantly decompacted chromatin for the 249kb stretch between the probes WI2-671I21 and WI2-1737E8 compared to wildtype, but that CN2 does not. The 517kb stretch of chromatin between WI2-671I21 and WI2-412A22 is significantly decompacted in both of the CdLS cell lines compared to wildtype. There is no evidence of change in the compaction of chromatin between WI2-671I21 and WI2-2372A13, a 750kb stretch, in CdLS compared to wildtype. At the 996kb separation between WI2-671I21 and WI2-3325E18, chromatin is significantly decompacted in CN1 compared to wildtype, however this is no greater than the general decompaction in chromatin as indicated by nuclear area, and there is no change in the chromatin compaction for this region in CN2 compared to wildtype (Figure 4.5C; Figure 4.6; Table 4.1; Table 4.2).

There does appear to be some decompaction of chromatin at 11q13 in CdLS compared to wildtype, however it is variable in two ways. Firstly, CN1 seems to show more chromatin decompaction than CN2 at this region, with CN1 giving more significant differences to W1 than CN2, and the increase in mean compared to W1 being much larger. This suggests variability between CdLS cell lines, perhaps explaining the variability in phenotype severity between patients. Secondly the chromatin appears to decompact more at the smaller separations in CdLS compared to wildtype, perhaps due to more noise at the larger separations which will have more flexibility than the smaller separations anyway, or perhaps this reflects the scale over which cohesin mediated structures have their effect. To understand this further, I considered another region with similar properties to 11q13.

Like 11q13 the epidermal differentiation complex (EDC) on chromosome 1 is a region that has also been studied for chromatin compaction in the lab (D. Sproul,

PhD Thesis, 2008), is also very open and gene-rich (Figure 4.7A). It has been shown that this region is located external to chromosome 1 in keratinocytes, where the genes are active, but close to the CT periphery in other cells, where the EDC is inactive (Williams *et al*, 2002). Later analysis of CTCF binding and genes misregulated in CdLS revealed that this region also has a high density of CTCF binding, 1554 sites/Mb (Barski *et al*, 2007), has a 14 cohesin binding sites across the 1Mb region studied and contains a small number of genes misregulated in CdLS LCLs (Liu *et al*, 2009).

Because this region has similar properties as 11q13, it was similarly studied using pairs of probes ~250kb, ~500kb, ~750kb and ~1Mb apart. In this case the anchor probe, WI2-2080I19, labelled in DIG-11-dUTP, was at the telomeric end of the region, compared to the four biotin-16-dUTP--labelled distance probes, WI2-2423D6 (262kb), WI2-553L9 (552kb), WI2-3057F14 (741kb) and WI2-784N16 (1055kb; Figure 4.7A; Figure 4.7B; Table 2.7).

At the EDC region, interphase separation increases with genomic distance, and there is a linear relationship between the r^2 and the genomic distance (Figure 4.7C). All of the datasets conform to a Rayleigh distribution, having a ratio of median to mean of ~0.9, and a ratio of standard deviation to mean of ~0.5 (Table 4.1).

The 262kb stretch of chromatin between WI2-2080I19 and WI2-2423D6 and the 741kb stretch between WI2-2080I19 and WI2-3057F14 at the EDC region are both significantly decompacted in both of the CdLS cell lines compared to wildtype. At 552kb separation (WI2-2080I19 and WI2-553L9) and the 1055 separation (WI2-2080I19 and WI2-784N16) chromatin is significantly decompacted in CN1 compared to wildtype, but this is no greater than the general decompaction in chromatin as indicated by nuclear area. Despite this there is no change in the chromatin compaction for this region in CN2 compared to wildtype (Figure 4.8; Table 4.1; Table 4.2).

Similarly to 11q13, the EDC region does show decompaction of chromatin in CdLS compared to wildtype, but, as with 11q13, this is highly variable. Again, the effect is stronger in CN1 compared to CN2. However, there is no indication that this may be related to the size of interval measured.

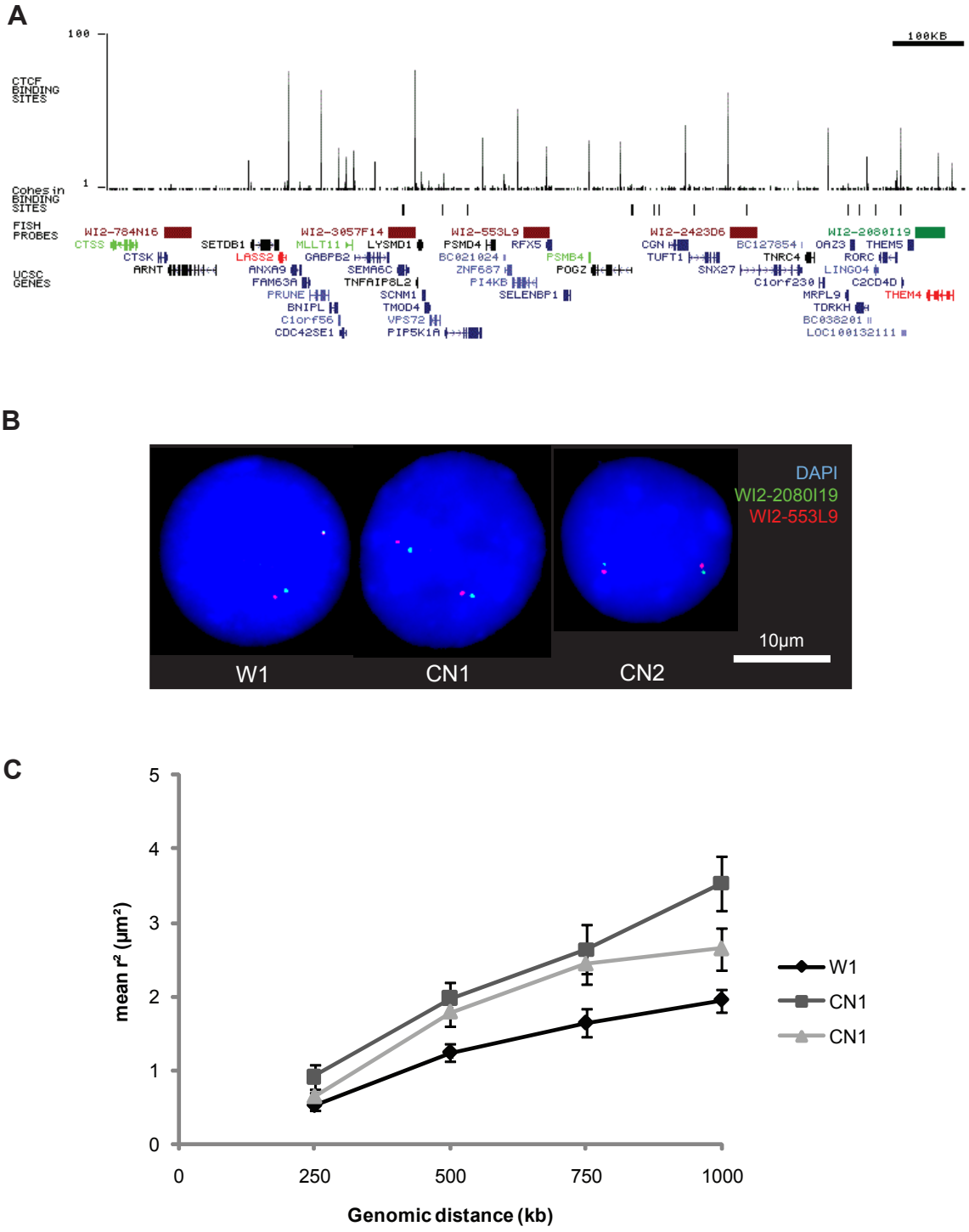


Figure 4.7 – Chromatin is decompacted the EDC in CdLS.
A, UCSC track of the human EDC (chr1:150,000,000-151,900,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, sample fluorescence microscopy images of nuclei hybridised to probes separated by ~500kb. **C**, The mean interphase separation squared is plotted against the genomic distance as measured in W1, CN1 and CN2 cell lines. The error bars represent the SEM.

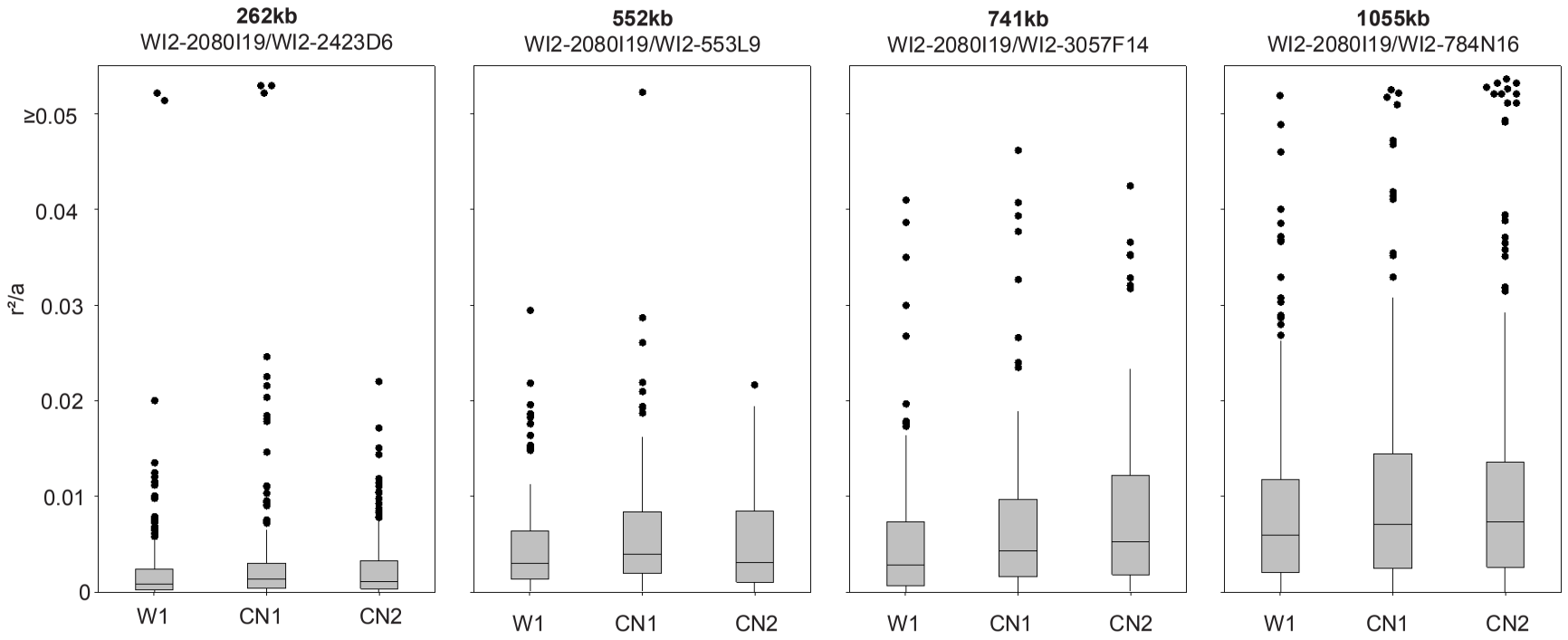


Figure 4.8 – Chromatin is decompacted the EDC in CdLS.
Boxplots showing the normalised interphase separation squared (r^2/a) of the region, at the four genomic distances in the three cell lines, as **Figure 3.2**.

4.3.2 Chromatin compaction at regions of lower CTCF binding density in CdLS cell lines

The mouse HoxD is another region that has been well studied in the Bickmore lab, and is found to decondense and loop out of its chromosome territory upon activation in mouse ES cell differentiation and embryogenesis (Morey *et al*, 2007). Chromatin compaction at this region in ES cells is shown to be mediated by polycomb repressive complexes (Eskeland *et al*, 2010). Because of this I decided to study this region, which is on human chromosome 2, in CdLS cell lines. The region around HoxD is relatively gene poor, containing few genes other than the HoxD cluster itself. The CTCF binding density of the region is fairly representative of the genome as a whole, at 883 sites/Mb (Barski *et al*, 2007) and there are seven cohesin binding sites within the 1Mb region studied (Liu *et* al, 2009; Figure 4.9A).

I was also interested in the HoxD locus for two other reasons. Firstly HoxD is known to be regulated by enhancers and as such is a region likely to be affected by a disorder that disrupts promoter/enhancer interactions. The global control region (GCR) is found 240kb upstream of the HoxD cluster and contains a number of enhancers that regulate HoxD expression in limbs and in the CNS (Figure 4.9A; Gérard *et al*, 1993; Spitz *et al*, 2003). Secondly the limb phenotype observed in CdLS (OMIM 122470) is reminiscent of the limb phenotype observed upon HoxD mutation (del Campo *et al*, 1999), suggesting that HoxD may be misregulated in CdLS. I had already studied the region by the time data came out showing that HoxD genes are not misregulated in CdLS LCLs (Liu *et al*, 2009), but it is more likely that HoxD genes are only misregulated in the limb bud cells and not LCLs. As developmental patterning genes, HoxD genes are not expressed in LCLs (Liu *et al*, 2009).

Like 11q13 and EDC, I chose separations of ~250kb, ~500kb, ~750kb and ~1Mb, but I also chose to study a separation of ~375kb as I hypothesised that this region would be more compact than 11q13 and the EDC so I wanted to have another small separation to find the more subtle changes at small separations. My DIG-11-dUTP-labelled anchor probe was WI2-3163O19 and my biotin-16-dUTP-labelled distance probes were WI2-1038D9 (253kb), WI2-2744M3 (396kb), WI2-1919C15 (491kb),

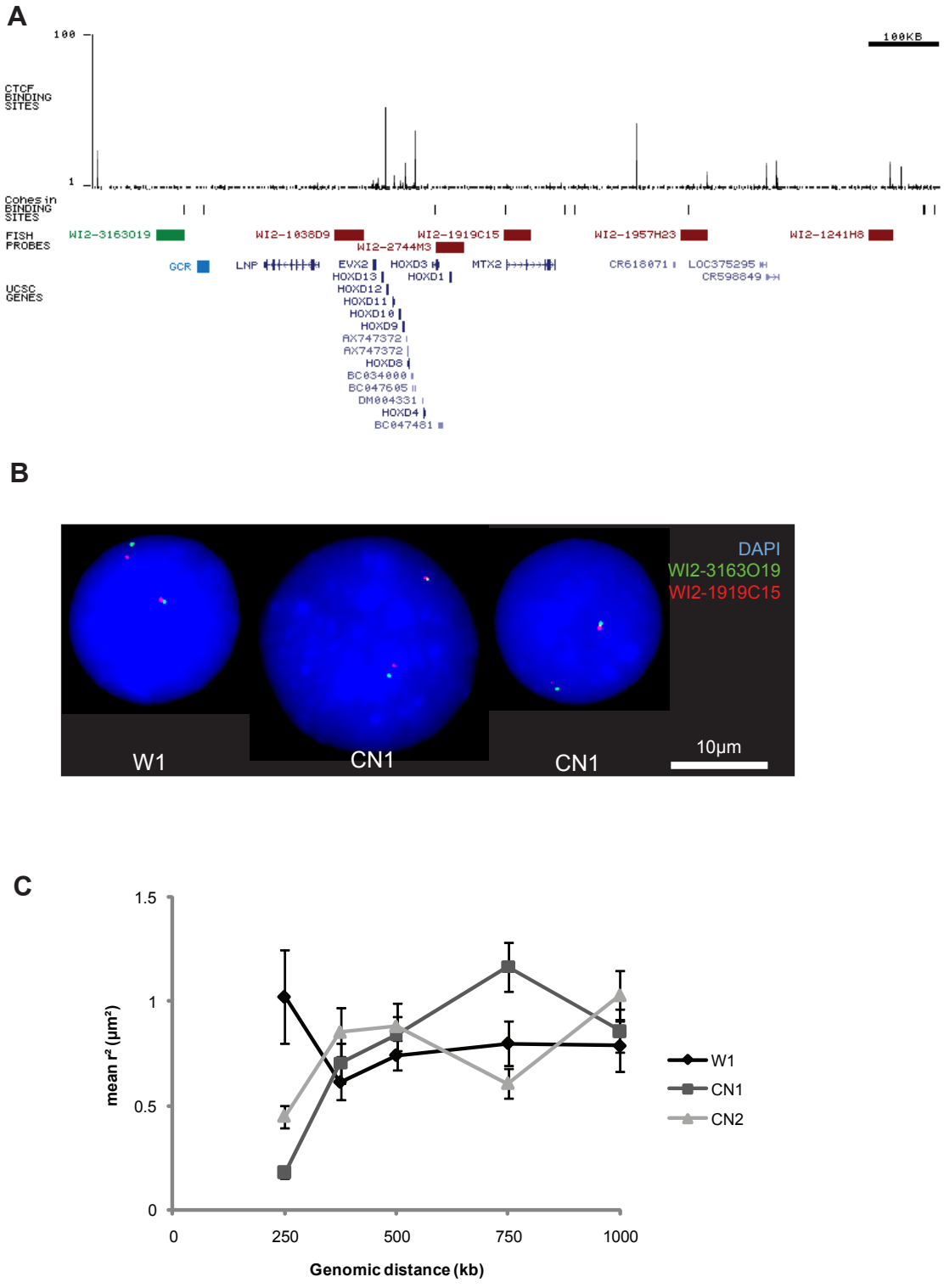


Figure 4.9 – Chromatin is not decompacted at HoxD in CdLS.
A, UCSC track of the human HoxD cluster (chr2:176,500,000-177,700,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. The position of the GCR is shown in turquoise. **B**, sample fluorescence microscopy images of nuclei hybridised to probes separated by ~500kb. **C**, The mean interphase separation squared is plotted against the genomic distance as measured in W1, CN1 and CN2 cell lines. The error bars represent the SEM.

WI2-1957H23 (742kb) and WI2-1241H8 (1007kb; Table 2.7; Figure 4.9A). The GCR is located in between my reference probe WI2-3163O19 and my 253kb probe WI2-1038D9.

This region is more condensed than both the EDC and 11q13 in all three cell lines: the mean and median r^2/a at the ~1Mb for the HoxD cluster is around half those of the EDC and 11q13 for all cell lines (Figure 4.6; Figure 4.8; Figure 4.10; Table 4.1; Table 4.2). This is because the HoxD cluster is compacted when it is inactive (Morey *et al*, 2007; Eskeland *et al*, 2010), as it is in LCLs.

There is some increase in r^2 with genomic distance at the HoxD locus, but there is deviation from a linear relationship (Figure 4.9C). Despite this, most datasets follow a Rayleigh distribution, with the ratio of median to mean of ~0.9 and a ratio of standard deviation to mean of ~0.5, except at the ~250kb separation, where the median is significantly lower than the mean, and the standard deviation is higher than the mean in W1 cells (Table 4.1). This suggests that there may be some specific interactions that take place within this region, that interfere with any direct correlation between r^2 and genomic distance and disrupt the random walk.

The data obtained with the 253kb separated probes (WI2-3163O19 and WI2-1038D9) suggests that there is a complex interaction that is disturbed by CdLS within this region, as CN1 appears to have condensed chromatin compared to wildtype whilst CN2 appears to have decondensed chromatin compared to wildtype. The distribution of r^2 values results in means are significantly higher than the medians for this separation, so whilst the mean r^2 for this region is significantly lower in CN2 than in W1 (Figure 4.9C), the distribution of data is such that most CN2 cells had higher r^2 values than W1 cells. The GCR is contained between these two probes and this locus may interact with other regions, disruption of which in CdLS could cause the complex phenotype observed here (Figure 4.8A; Barski *et al*, 2007). There is no evidence of change in the compaction of chromatin at the 396kb (WI2-3163O19 and WI2-2744M3) and 491kb (WI2-3163O19 and WI2-1919C15). At the 742kb separation (WI2-3163O19 and WI2-1957H23) chromatin is significantly decompacted in CN1 compared to wildtype, however this is no greater than the general decompaction in chromatin as indicated by nuclear area. However

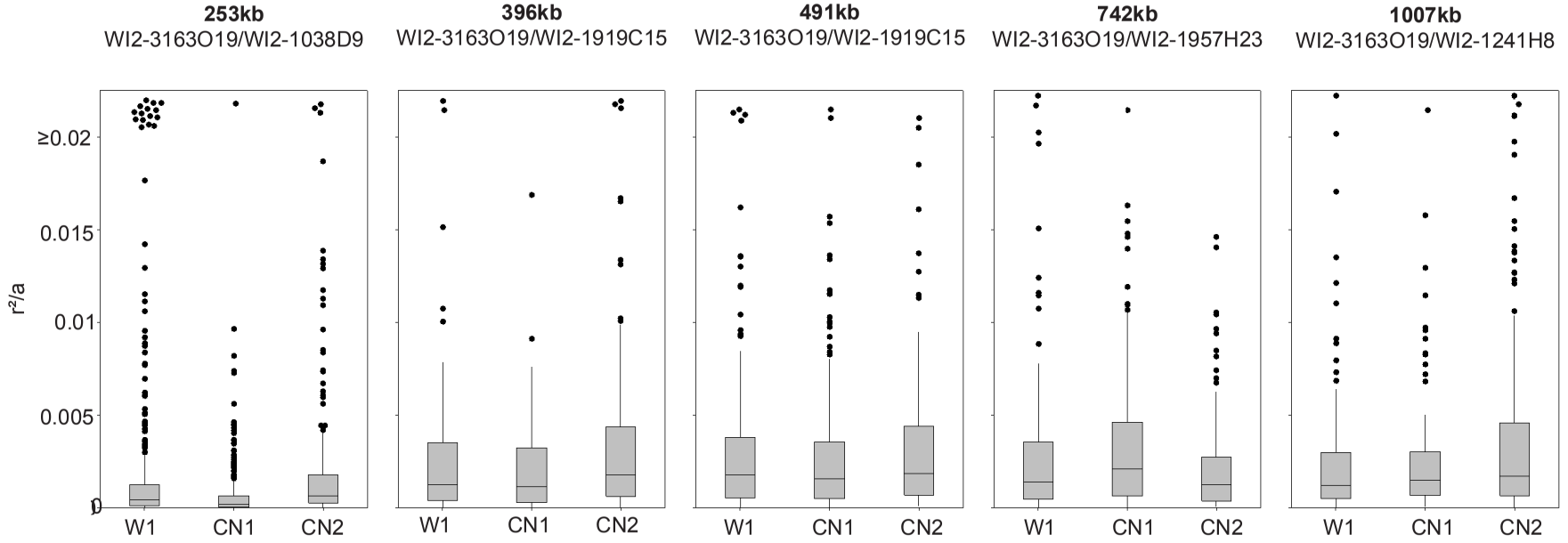


Figure 4.10 – Chromatin is not decompacted at HoxD in CdLS.
Boxplots showing the normalised interphase separation squared (r^2/a) of the region, at the five genomic distances in the three cell lines, as **Figure 3.2**.

there is no change in the chromatin compaction for this region in CN2 compared to wildtype. There is no change in the compaction of chromatin between WI2-3163O19 and WI2-1241H8, a stretch 1007kb long, in CN1 compared to W1, but some decompaction of CN2 (Figure 4.10; Table 4.1; Table 4.2).

The link between CdLS and chromatin compaction is not clear at HoxD, a region with a CTCF binding density not dissimilar to the average genomic density. The HoxD gene cluster is tightly regulated and if cohesin is involved in this regulation, it is likely that its interaction here is complex. Cohesin may be linked to the interactions of the GCR with other regions, causing complex changes in CdLS at the 253kb separation which contains the GCR. Even though the GCR is also included in the larger separations, the effect of having more chromatin within the region may dilute out the effects of the GCR. If this is the case, then FISH may be too crude an assay to delineate these interactions.

I wanted to find a gene desert to contrast with the three regions I had already studied, which all contained some genes and had some CTCF and cohesin binding, so I chose a region at 18q22, which contained a 500kb gene-free region (Figure 4.11A). This region also has a very low CTCF binding density (788 sites/Mb; Barski *et al*, 2007), no cohesin binding sites identified by ChIP-chip (Liu *et al*, 2009) and the only two genes in the vicinity are not misregulated in CdLS (Liu *et al*, 2009).

In order to keep all my probes within the gene desert, I only used separations of ~250kb, ~375kb and ~500kb. The probes used were WI2-1702P7 as my anchor probe, WI2-502C21 (239kb), WI2-1735D16 (376kb) and WI2-1708L7 for (486kb; Table 2.7; Figure 4.11A).

At 18q22, interphase separation increases with genomic distance, and there is an approximately linear relationship between the r^2 and the genomic distance in all three cell lines (Figure 4.11; Table 4.1). Like HoxD, this region is more compacted than 11q13 and the EDC, due to this regions being a gene desert, which tend to be more compact than gene-rich regions (Gilbert *et al*, 2004).

For neither the 239kb stretch (WI2-1702P7 and WI2-502C21), nor the 376kb stretch (WI2-1702P7 and WI2-1735D16), is there any change in chromatin compaction in either CdLS cell line. Between the 486kb separated probes WI2-

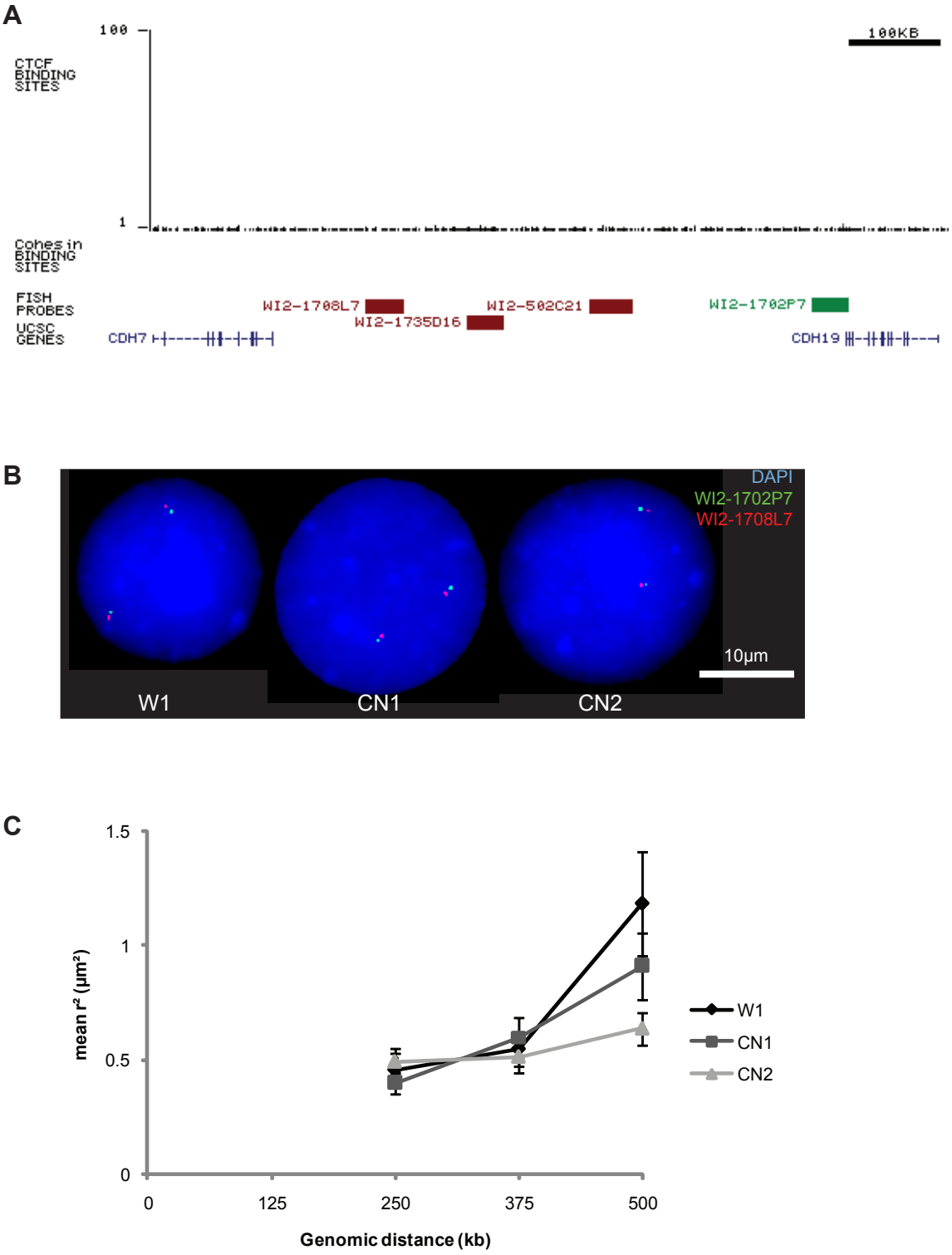


Figure 4.11 – Chromatin is not decompacted at 18q22 in CdLS.
A, UCSC track of 18q22 (chr18:63,450,000-64,300,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, sample fluorescence microscopy images of nuclei hybridised to probes separated by ~500kb. **C**, The mean interphase separation squared is plotted against the genomic distance as measured in W1, CN1 and CN2 cell lines. The error bars represent the SEM.

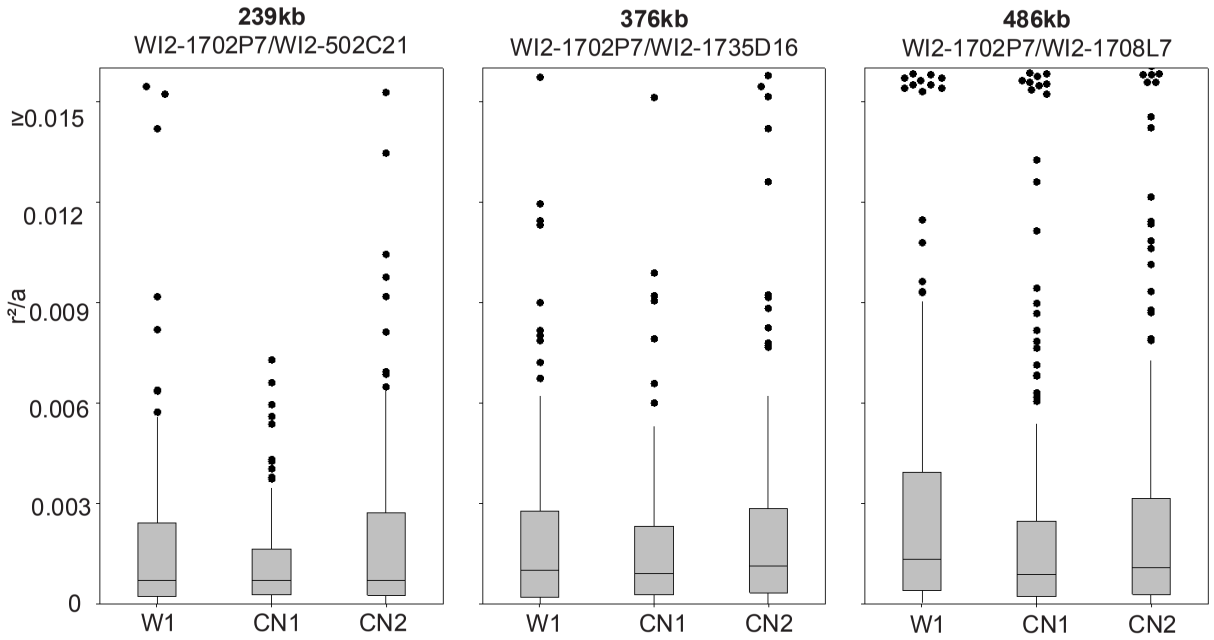


Figure 4.12 – Chromatin is not decompacted at 18q22 in CdLS.
Boxplots showing the normalised interphase separation squared (r^2/a) of the region, at the three genomic distances in the three cell lines, as **Figure 3.2**.

1702P7 and WI2-1708L7, there is evidence that the chromatin is more compacted in CdLS than in wildtype (Figure 4.12; Table 4.1; Table 4.2).

The absence of any CTCF- or cohesion-binding sites in 18q22 makes it unsurprising that at the ~250kb and ~375kb separations there is no change in chromatin compaction. However the apparent compaction of chromatin in CdLS at the ~500kb separation is an unexpected result.

4.3.3 How CdLS affects different genomic regions

The effects of CdLS differ at different regions. At both the EDC and 11q13 the chromatin is generally less compact in CdLS cells compared to wildtype, whereas at 18q22 the chromatin is more compact, and at HoxD this relationship is more complex. To understand this, we must think in terms of chromosome topology, and not compaction of chromatin itself. If cohesin is indeed able to regulate gene expression by tethering chromatin loops together, it is this that may cause the difference in interphase distances observed, and not compaction of the 30nm fibres.

Both the EDC and 11q13 are gene-rich regions with dense cohesin binding (Figure 4.5A; Figure 4.7A). A number of genes within these two regions are likely to be regulated by cohesin mediated looping as a number of genes in each region are misregulated in CdLS LCLs (Liu *et al*, 2009), and loops within these two regions will result in an apparent compaction of the chromatin. Hence reduction in cohesin in CdLS will result in loss of these loops and an apparent chromatin compaction.

Genes of the HoxD cluster are unlikely to be expressed in LCLs, however the tight and complex regulation of these genes means that this region is still affected in CdLS. The cohesin binding sites within this region, particularly those associated with the GCR (Figure 4.9A), may hold the HoxD in a complex structure that is disturbed in CdLS. A 5C study of this region could delineate this structure to further understand how HoxD is affected by CdLS.

The gene-poor, cohesin-poor 18q22 region is more compacted in CdLS compared to wildtype. However on either side of the region studied, 600kb telomeric and 400kb centromeric, are two cohesin binding sites and a cohesin

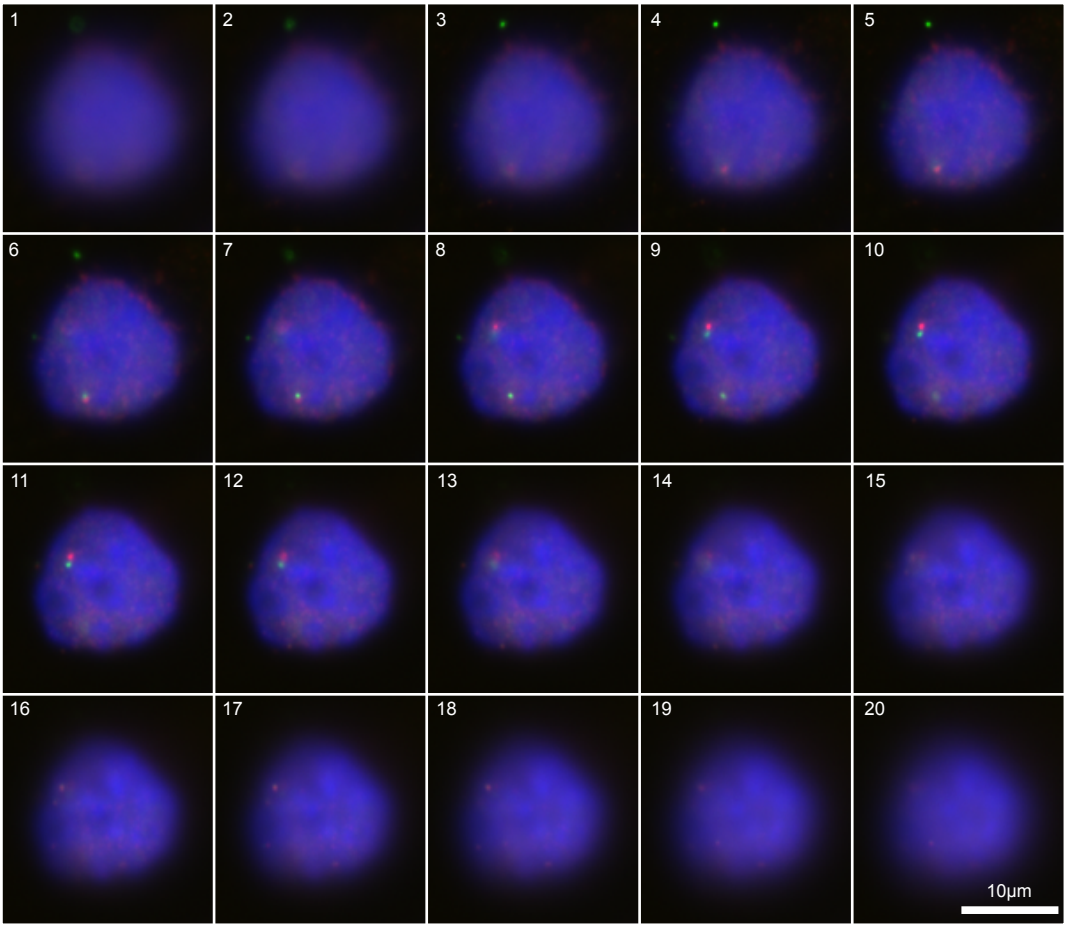
binding site cluster 1.6Mb centromeric to the region, identified by ChIP (Liu *et al*, 2009). It is plausible that the interactions of these sites with other genomic loci leads to a stretching of the 18q22 region in wildtype cells, which is abrogated in CdLS, giving compaction of this region.

Chromatin compaction in CdLS may also be affected by chromatin modifications. NIPBL has been shown to recruit HDAC1 and HDAC3, leading to repression of reporter genes; CdLS mutations result in loss of this repression (Jahnke *et al*, 2008). Presumably, this effect could lead to the conversion of all chromatin to an active and decompacted state, with an increase in nuclear size. However, the decompaction I have observed in CdLS is not universal, suggesting that this is not the mechanism by which cohesin regulates cohesin structure. If any changes in histone acetylation are occurring in CdLS, instead of having a knock-on effect on other histone modifications and DNA methylation that can influence chromatin compaction, these other modifications may maintain correct compaction. Indeed, there is no alteration in the DNA methylation state of autosomal chromatin in CdLS (Liu *et al*, 2010).

4.4 Confirmation of 2D FISH data by 3D FISH

Whilst 2D FISH has been shown to be a robust method for determining compaction of chromatin (Morey *et* al, 2007; Eskeland *et al*, 2010), it is important to replicate the data in 3D nuclei, to confirm that the changes observed in chromatin are not an artefact of the fixation process. The method for 3D FISH differs from that for 2D FISH in the method used to fix the nuclei; in the latter MAA fixation dehydrates the nucleus and extracts many proteins. In 3D FISH it is important to maintain the 3D structure of the nucleus so live cells are adhered to the slides by cytospinning then fixed with pFA (Section 2.8). To image 3D nuclei, I used a moving objective to image the nucleus as sections across the *z*-axis at known intervals. The technology was not available for me to construct 3D maps of the nuclei, so I could not directly measure the 3D distance between FISH probes. Instead I measured the *xy* distance from a maximal projection of the stacks and counted the number of stacks between the probes to calculate the *z* distance; I could then calculate the *xyz* distance (Figure

A



B

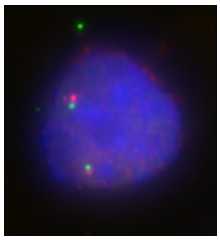


Figure 4.13 - Analysis of 3D FISH image stacks.
A, A mosiac of the 20 image stacks taken for 3D FISH analysis, at 0.25µm *z* axis intervals. **B**, A merged 2D image made from a maximal projection of the 3D stack. To analyse 3D FISH images the *xy* distance between probes is calculated from this image using the 2D FISH distance script (Figure 2.3); in this image, this gives a distance of 0.36µm for the bottom probe pair. The *z* distance is calculated by identifying the images in the 3D stack in which the probe can be seen the most clearly, and are the most in focus; in this image stack the green signal in the bottom probe pair is most in focus in image 8, and the red signal of the bottom probe pair is most in focus in image 6, giving a *z* distance of 0.5µm. The *xyz* distance can now be calculated from these two values using Pythagoras Theorum. Distance = $\sqrt{(0.36^2 + 0.5^2)} = 0.62\mu m$.

4.13). This method of calculation of physical distances means that 3D FISH is more labour intensive than 2D and gains less accurate results, as it relies upon human judgement and less detailed measurement. Therefore 3D FISH is mostly suitable for confirmation of 2D FISH data, and not for collection of large-scale datasets, and 3D FISH was only carried out using two pairs of probes.

4.4.1 Chromatin is decompacted at regions of high CTCF binding density in CdLS

I decided to use 3D FISH to confirm the data I found at 11q13 with a ~500kb separation, as at this region and separation both CN1 and CN2 CdLS cell lines showed a significant decompaction of chromatin compared to W1 control cells. I used the probes WI2-671I21 and WI2-412A22 that I had used earlier to study this region in 3D (Figure 4.14A).

As with the ~500kb separation studied at 11q13 in 2D, CN1 has significantly decompacted chromatin compared to W1, however unlike the 2D data, CN2 does not (Figure 4.14B; Table 4.3). Furthermore, across the region in the 2D study, CN2 did not have decompacted chromatin as consistently as CN1. The distances measured in 3D are smaller than those measured in 2D (Figure 4.14B); this is due to differences in the fixing method and have been previously observed in similar experiments (Morey *et al*, 2007; Eskeland *et al*, 2010). These data, therefore, do support the data collected in 2D, and show that the 2D data is reflective of the chromatin compaction in 3D fixed cells.

4.4.2 Chromatin maintains its compaction at regions of low CTCF binding density in CdLS

I also carried out 3D FISH at 18q22 using the ~500kb separation as this region showed an opposite effect to 11q13 in CdLS, with both CN1 and CN2 having more compact chromatin compared to W1 at this region and separation. I also felt that using ~500kb separations at both regions would make for more effective comparison

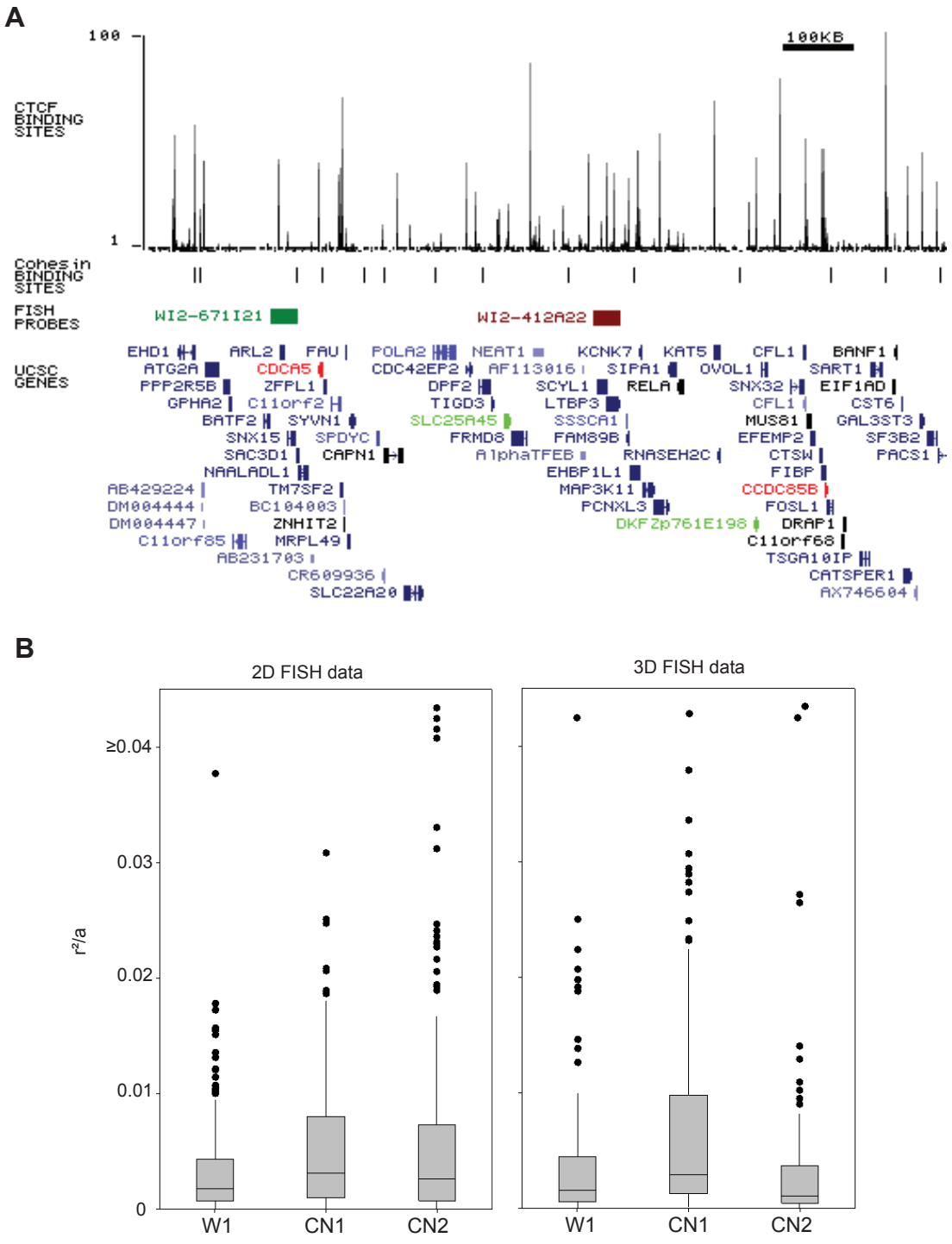


Figure 4.14 – Chromatin decompaction at 11q13 in CdLS can be confirmed in 3D.
A, UCSC track of 11q13 (chr11:64,500,000-65,900,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a) of the region as measured in 2D and 3D, as **Figure 3.2**.

Table 4.3 - 3D FISH statistics

The statistics of every r^2 and r^2/a dataset studied by 3D FISH. Significant differences are highlighted in grey.

r^2

Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2 (μm^2)	Median r^2 (μm^2)	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?	median r/ mean r	stdev r/ mean r
11q13	WI2-671I21	WI2-412A22	517	W1	0.53	0.21	0.000			0.74	0.78
				CN1	1.04	0.39		0.000	no	0.75	0.76
				CN2	0.55	0.16		0.047	yes	0.68	1.02
18q22	WI2-1702P7	WI2-1708L7	486	W1	0.25	0.11	0.128			0.79	0.73
				CN1	0.32	0.11		0.234	yes	0.64	1.44
				CN2	0.25	0.09		0.079	yes	0.84	0.97

r^2/a

Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2/a	Median r^2/a	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?
11q13	WI2-671I21	WI2-412A22	517	W1	0.0043	0.0016	0.000		
				CN1	0.0078	0.0029		0.002	no
				CN2	0.0046	0.0011		0.137	yes
18q22	WI2-1702P7	WI2-1708L7	486	W1	0.0021	0.0008	0.226		
				CN1	0.0021	0.0006		0.087	yes
				CN2	0.0021	0.0006		0.366	yes

of data. I used the probes WI2-1702P7 and WI2-1708L7 for this study as before (Figure 4.15A).

There is no evidence of a statistically significant change in the compaction of chromatin between WI2-1702P7 and WI2-1708L7 in CdLS compared to wildtype when studied in 3D. Visually, it would appear that this region is more compact in CN1 than in W1 and CN2, however this is not statistically significant (Figure 4.15B; Table 4.3). These data are more similar to the data collected for the ~250kb and ~375kb separation in 2D, suggesting that the apparent decompacted chromatin at ~500kb in W1 in 2D was an anomaly, however the rest of the data collected in 2D is reflective of living cells.

4.5 The effect of CdLS on chromatin is not dependant on the cell cycle

Given the importance of cohesin in the cell cycle, it was important to show that the effects I observed were not specific to one stage of the cell cycle. To this end, I fractioned cells based on size using centrifugal elutriation, as cells are larger in later stages of the cell cycle. Centrifugal elutriation combines a pump with a centrifuge, pumping cells in the opposite direction to the centrifugal force. As the pump speed increases, larger cells are able to escape the centrifugal force and be pumped out, so by moving up the pump speed in measured increments, fractions of cells of different sizes can be collected (Section 2.6.4). I was then able to analyse the DNA content of the cells using FACS, and choose the fractions that best represented each stage of the cell cycle.

As seen in the previous chapter, we already know that in a mixed population of cells, CN1 and CN2 will have a larger number of cells in the two growth phases of the cell cycle than W1. By centrifugal elutriation, I succeeded in producing populations that were enriched for G_1 cells, with all cell lines with over 70% of cells with 2n chromosomes in the G_1 populations as judged by FACS. I was able to produce G_2 populations that were enriched for cells with 4n chromosomes, increasing the proportion of cells in G_2 from below 10% to over 40% in CN1 and

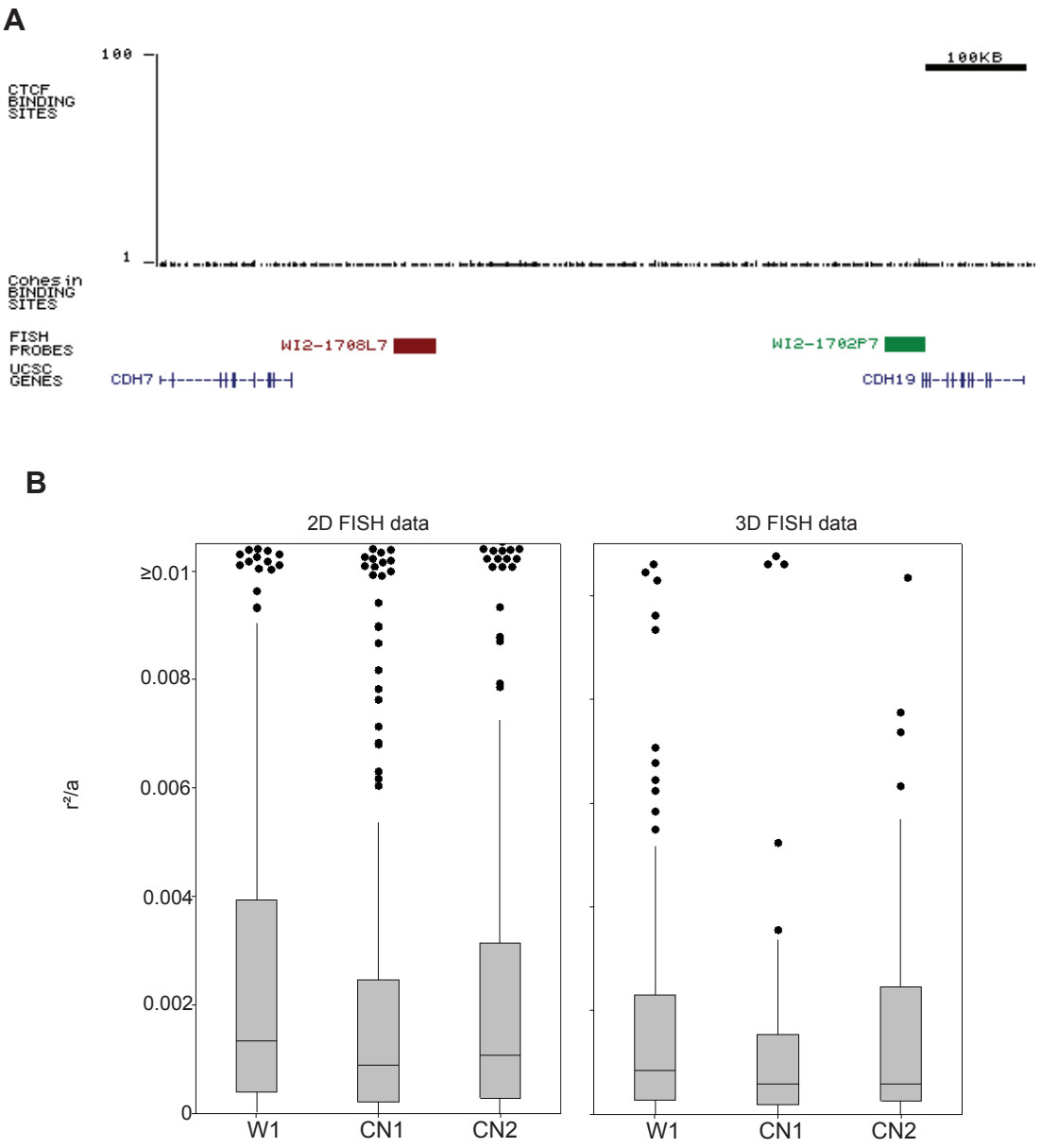


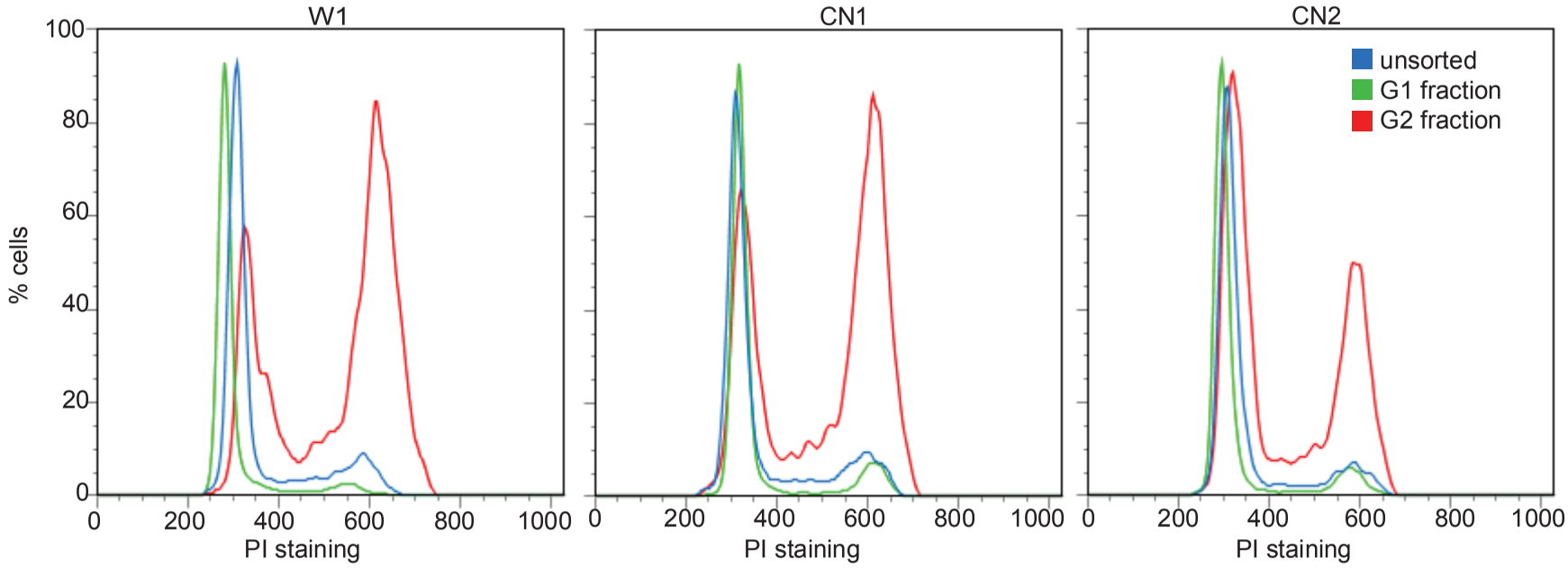
Figure 4.15 – The lack of chromatin decompaction at 18q22 in CdLS can be confirmed in 3D. **A**, UCSC track of 18q22 (chr18:63,450,000-64,300,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a) as measured in 2D and 3D, as **Figure 3.2**.

W1, and over 30% in CN2 (Figure 4.16). Although my G_2 populations do not contain a majority of G_2 cells, they contain a sufficient proportion of G_2 cells to show a difference to a mixed population.

4.5.1 Chromatin is decompacted at regions of high CTCF binding density in CdLS throughout the cell cycle

I again used the 500kb separation at the 11q13 region for this analysis, using probes WI2-671I21 and WI2-412A22 (Figure 4.17A). In the G_1 population, chromatin is significantly decompacted in CN1 compared to wildtype at this region, however this is no greater than the general decompaction in chromatin as indicated by nuclear area, and there is no change in the chromatin compaction for this region in CN2 compared to wildtype (Figure 4.17B; Table 4.4). The G_2 population has a similar effect, the chromatin of CN2 is unaffected but the chromatin of CN1 appears to decondense, but this effect is erased upon normalisation to nuclear area (Figure 4.17C; Table 4.4). Comparing the G_1 and G_2 datasets, we can see that r^2 measurements are much higher in G_2 than in G_1, reflecting the decompaction of chromatin through the cell cycle, but r^2/a measurements do not differ between G_1 and G_2 suggesting that this is a general decompaction of all chromatin through the cell cycle, not of specific regions (Table 4.4; Figure 4.17B).

The cell cycle specific data at 11q13 does not completely support the data in the mixed population, where chromatin was shown to decompact in CdLS, as the chromatin does not decompact more than the general decompaction of the nucleus in either G_1 or G_2. However I have already observed that the phenotypes in CdLS are often inconsistent and variable. Also, the G_1 and G_2 populations behave the same as each other, showing that any effect we observe is not restricted to G_2, when cohesin is involved in sister chromatid cohesion.



	% cells in		
	G1	**S-phase**	**G2**
unsorted	45.6	34.1	4.4
G1 fraction	77.2	14.9	4.6
G2 fraction	20.5	29.0	46.9

	% cells in		
	G1	**S-phase**	**G2**
unsorted	62.0	21.2	9.7
G1 fraction	74.0	12.8	10.2
G2 fraction	24.5	22.5	47.6

	% cells in		
	G1	**S-phase**	**G2**
unsorted	70.8	15.6	7.9
G1 fraction	75.3	16.3	6.4
G2 fraction	46.6	16.9	30.8

Figure 4.16 – FACS profiles of cells sorted by Centrifugal Elutriation.
Histograms of DNA content of unsorted W1, CN1 and CN2 cells, the G1 fraction and the G2 fraction sorted by centrifugal elutriation. The proportion of each fraction in each cell cycle stage is shown in the tables below.

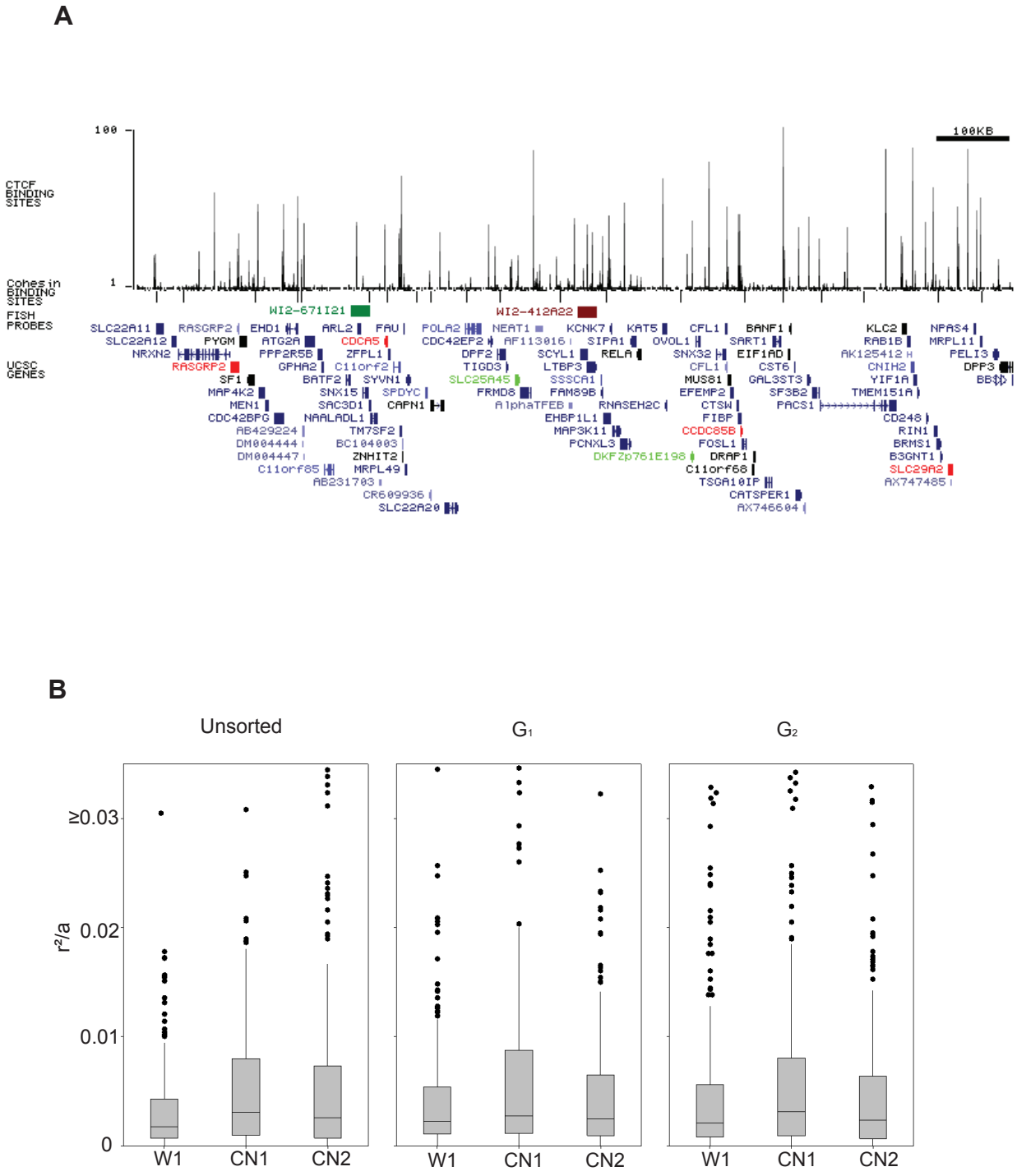


Figure 4.17 - Chromatin decompaction at 11q13 in CdLS occurs throughout the cell cycle. **A**, UCSC track of 11q13 (chr11:64,500,000-65,900,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a) in the unsorted cells and the G1 and G2 fractions, as **Figure 3.2**.

Table 4.4 - Cell cycle FISH statistics

The statistics of every cell cycle specific r^2 and r^2/a dataset studied by 2D FISH. Significant differences are highlighted in grey.

r^2

Cell cycle stage	Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2 (µm²)	Median r^2 (µm²)	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?	median r/ mean r	stdev r/ mean r
G1	11q13	WI2-671I21	WI2-412A22	517	W1	0.35	0.19	0.001			0.87	0.61
					CN1	0.63	0.30		0.000	no	0.82	0.64
					CN2	0.39	0.21		0.753	yes	0.87	0.63
	18q22	WI2-1702P7	WI2-1708L7	486	W1	2.76	0.22	0.089			0.91	0.65
					CN1	0.74	0.13		0.002	yes	0.70	1.31
					CN2	0.41	0.18		0.431	yes	0.835941W13	0.76
G2	11q13	WI2-671I21	WI2-412A22	517	W1	0.75	0.27	0.000			0.76	0.77
					CN1	1.49	0.78		0.000	no	0.89	0.72
					CN2	0.97	0.55		0.010	yes	0.92	0.70
	18q22	WI2-1702P7	WI2-1708L7	486	W1	0.18	0.07	0.000			0.79	0.72
					CN1	0.55	0.18		0.000	no	0.73	0.82
					CN2	0.59	0.26		0.000	no	0.84	0.80

r^2/a

Cell cycle stage	Region	Probe 1	Probe 2	Separation (kb)	Cell line	Mean r^2/a	Median r^2/a	Group ANOVA p-value	MW to W1 p-value	Tukey's CI overlap W1?
G1	11q13	WI2-671I21	WI2-412A22	517	W1	0.0042	0.0022	0.334		
					CN1	0.0057	0.0028		0.053	yes
					CN2	0.0047	0.0025		0.781	yes
	18q22	WI2-1702P7	WI2-1708L7	486	W1	2.3897	0.0022	0.050		
					CN1	0.0065	0.0013		0.007	yes
					CN2	0.0047	0.0019		0.672	yes
G2	11q13	WI2-671I21	WI2-412A22	517	W1	0.0047	0.0021	0.409		
					CN1	0.0064	0.0031		0.080	yes
					CN2	0.0051	0.0024		0.800	yes
	18q22	WI2-1702P7	WI2-1708L7	486	W1	0.0019	0.0007	0.022		
					CN1	0.0028	0.0011		0.019	yes
					CN2	0.0031	0.0012		0.001	no

4.5.2 Chromatin maintains its compaction at regions of low CTCF binding density in CdLS throughout the cell cycle

I again used the 500kb separation at the 18q22 region for this analysis, using probes WI2-1702P7 and WI2-1708L7 (Figure 4.18A).

In the G_1 population there is some evidence that CN1 may have condensed chromatin at 18q22 compared to W1, but CN2 does not (Figure 4.18B; Table 4.4). Surprisingly, both CdLS cell lines have decondensed chromatin compared to wildtype at 18q22 in G_2 (Figure 4.18C; Table 4.4). In all three cell lines, the r^2 measurements in the G_2 fraction are significantly smaller than those in G_1 fraction, whilst the r^2/a measurements decrease in G_2 compared to G_1 in W1, but there is little difference between G_1 and G_2 r^2/a measurements in CN1 and CN2 (Table 4.5; Figure 4.18).

I previously suggested that chromatin at 18q22 may be stretched out in wildtype cells by the interactions of two cohesin binding sites that flank the region. These data suggest that upon DNA replication the effect is lost in wildtype cells, perhaps due to altered gene expression through the cell cycle, and the region compacts. In CdLS there is no stretching in G_1 and no compaction in G_2, giving consistent compaction of this region through the cell cycle.

4.6 The effect on chromatin compaction does not occur with all CdLS cell lines

Having conducted extensive analysis of chromatin compaction with the CdLS cell lines CN1 and CN2, compared to the wildtype W1, I then began analysis on the new cell lines from Philadelphia, the wildtype lines WP1, WP2 and WP3 and the CdLS lines, CP1, CP2, CP3 and CP4. I decided to analyse just the 500kb separations at 11q13 and 18q22, as I had done with 3D FISH and with the cell cycle fractions, using WI2-671I21 and WI2-412A22 at 11q13 (Figure 4.19A) and WI2-1702P7 and WI2-1708L7at 18q22 (Figure 4.20A).

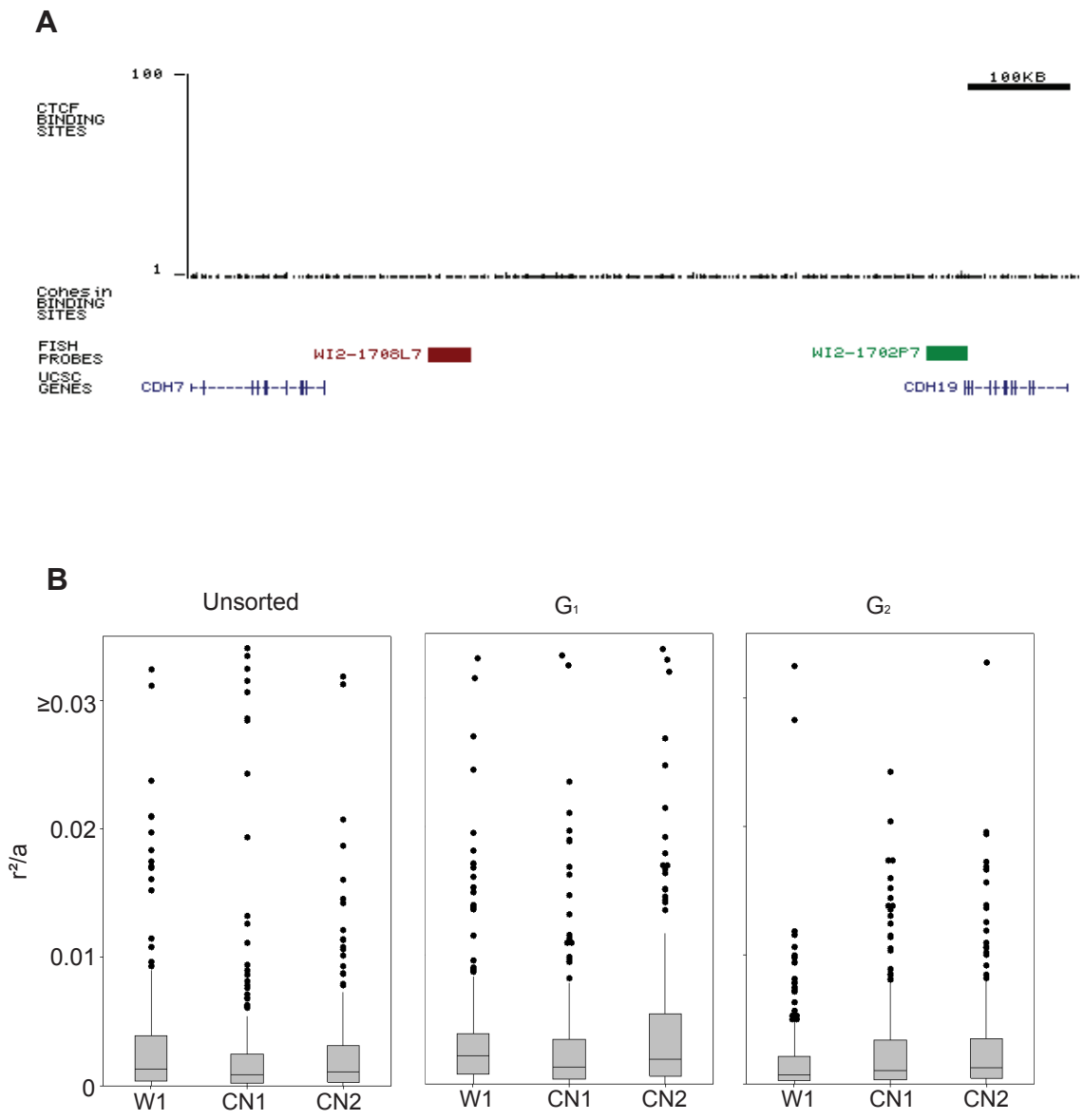


Figure 4.18 – The lack of chromatin decompaction at 18q22 in CdLS occurs throughout the cell cycle.
A, UCSC track of 18q22 (chr18:63,450,000-64,300,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a) in the unsorted cells and the G1 and G2 fractions, as **Figure 3.2**.

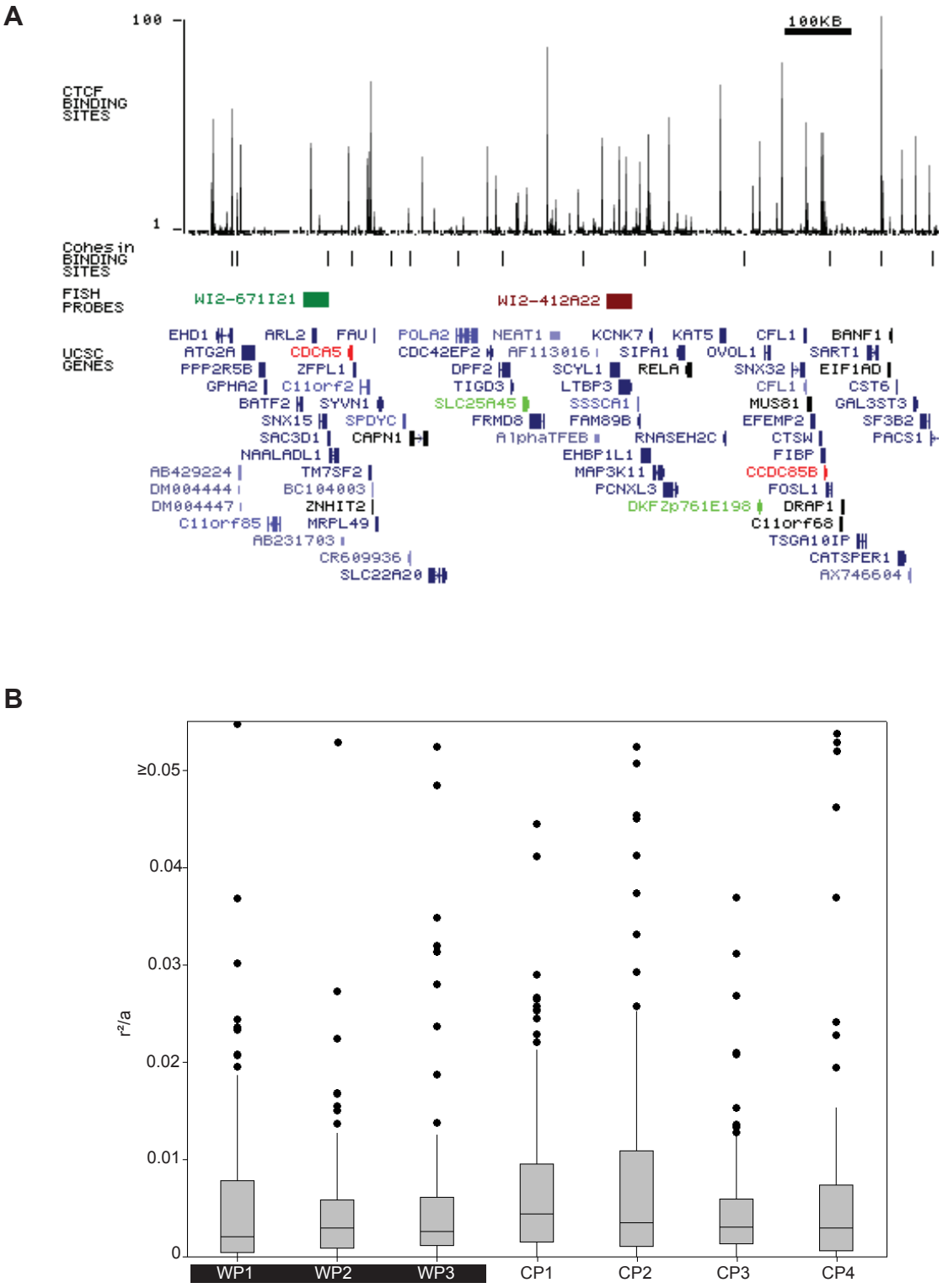


Figure 4.19 – Chromatin is not decompacted at 11q13 in all CdLS cell lines.
A, UCSC track of 11q13 (chr11:64,500,000-65,900,000; GRCh37/h19 assembly; Kent et al, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a), as **Figure 3.2**. Wildtype cell lines are highlighted in black.

Table 4.5 - 11q13 statistics

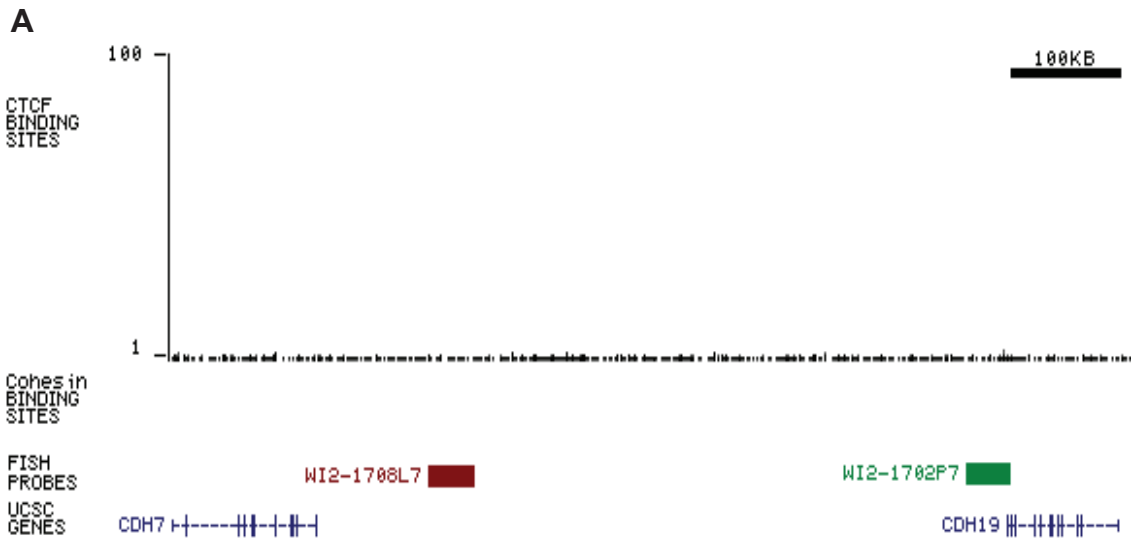
The statistics of every r² and r²/a dataset studied at 11q13 by 2D FISH with the Philadelphia cell lines, with pairwise statistical comparisons. Statistical differences are highlighted in grey.

r²

			WP1	WP2	WP3	CP1	CP2	CP3	CP4
mean (µm²)			1.14	1.09	1.34	1.86	1.67	1.35	1.34
median (µm²)			0.42	0.63	0.68	1.11	0.73	0.89	0.66
Compared to	WP1	MW	1	0.079	0.021	0.000	0.006	0.003	0.129
		Tukey's CI		yes	no	no	no	no	yes
	WP2	MW	0.079	1	0.638	0.001	0.152	0.077	0.970
		Tukey's CI	yes		yes	yes	yes	yes	yes
	WP3	MW	0.021	0.638	1	0.005	0.434	0.204	0.643
		Tukey's CI	no	yes		yes	yes	yes	yes
	CP1	MW	0.000	0.001	0.005	1	0.136	0.068	0.002
		Tukey's CI	no	yes	yes		yes	yes	yes
	CP2	MW	0.006	0.152	0.434	0.136	1	0.996	0.162
		Tukey's CI	no	yes	yes	yes		yes	yes
	CP3	MW	0.003	0.077	0.204	0.068	0.996	1	0.112
		Tukey's CI	no	yes	yes	yes	yes		yes
	CP4	MW	0.129	0.970	0.643	0.002	0.162	0.112	1
		Tukey's CI	yes	yes	yes	yes	yes	yes	

r²/a

			WP1	WP2	WP3	CP1	CP2	CP3	CP4
mean			0.0057	0.0047	0.0056	0.0074	0.0082	0.0049	0.0069
median			0.0021	0.0029	0.0028	0.0044	0.0036	0.0031	0.0029
Compared to	WP1	MW	1	0.431	0.177	0.002	0.015	0.161	0.322
		Tukey's CI		yes	yes	yes	yes	yes	yes
	WP2	MW	0.431	1	0.544	0.002	0.042	0.400	0.713
		Tukey's CI	yes		yes	yes	yes	yes	yes
	WP3	MW	0.177	0.544	1	0.012	0.156	0.750	0.937
		Tukey's CI	yes	yes		yes	yes	yes	yes
	CP1	MW	0.002	0.002	0.012	1	0.534	0.016	0.018
		Tukey's CI	yes	yes	yes		yes	yes	yes
	CP2	MW	0.015	0.042	0.156	0.534	1	0.185	0.133
		Tukey's CI	yes	yes	yes	yes		yes	yes
	CP3	MW	0.161	0.400	0.750	0.016	0.185	1	0.804
		Tukey's CI	yes	yes	yes	yes	yes		yes
	CP4	MW	0.322	0.713	0.937	0.018	0.133	0.804	1
		Tukey's CI	yes	yes	yes	yes	yes	yes	



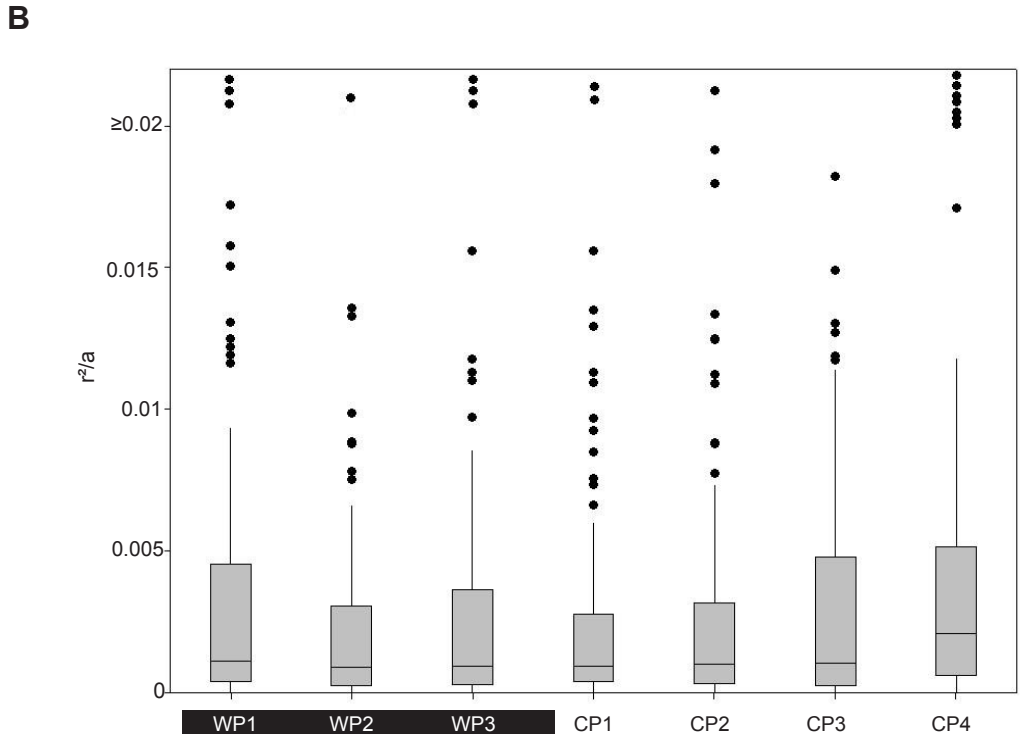


Figure 4.20 - Chromatin is not decompacted at 18q22 in CdLS cell lines.
A, UCSC track of 18q22 (chr18:63,450,000-64,300,000; GRCh37/h19 assembly; Kent et al, 2002; http://genome.ucsc.edu/) as **Figure 4.5**. **B**, Boxplots showing the normalised interphase separation squared (r^2/a), as **Figure 3.2**. Wildtype cell lines are highlighted in black.

Table 4.6 - 18q22 statistics

The statistics of every r^2 and r^2/a dataset studied at 18q22 by 2D FISH with the Philadelphia cell lines, with pairwise statistical comparisons. Statistical differences are highlighted in grey.

r^2

			WP1	WP2	WP3	CP1	CP2	CP3	CP4
mean (µm²)			0.77	0.56	0.46	0.41	0.70	0.90	0.98
median (µm²)			0.19	0.26	0.14	0.20	0.22	0.21	0.56
Compared to	WP1	MW	1	0.971	0.027	0.265	0.954	0.954	0.001
		Tukey's CI		yes	yes	yes	yes	yes	no
	WP2	MW	0.971	1	0.025	0.219	0.861	0.941	0.001
		Tukey's CI	yes		yes	yes	yes	yes	no
	WP3	MW	0.027	0.025	1	0.171	0.018	0.029	0.000
		Tukey's CI	yes	yes		yes	yes	yes	no
	CP1	MW	0.265	0.219	0.171	1	0.211	0.275	0.000
		Tukey's CI	yes	yes	yes		yes	yes	no
	CP2	MW	0.954	0.861	0.018	0.211	1	0.961	0.002
		Tukey's CI	yes	yes	yes	yes		yes	yes
	CP3	MW	0.954	0.941	0.029	0.275	0.961	1	0.003
		Tukey's CI	yes	yes	yes	yes	yes		no
	CP4	MW	0.001	0.001	0.000	0.000	0.002	0.003	1
		Tukey's CI	no	yes	no	no	yes	no	

r^2/a

			WP1	WP2	WP3	CP1	CP2	CP3	CP4
mean			0.0049	0.0024	0.0033	0.0023	0.0032	0.0049	0.0040
median			0.0012	0.0009	0.0009	0.0009	0.0010	0.0010	0.0022
Compared to	WP1	MW	1	0.065	0.207	0.100	0.329	0.380	0.177
		Tukey's CI		yes	yes	yes	yes	yes	yes
	WP2	MW	0.065	1	0.591	0.849	0.390	0.411	0.001
		Tukey's CI	yes		yes	yes	yes	yes	yes
	WP3	MW	0.207	0.591	1	0.783	0.689	0.761	0.012
		Tukey's CI	yes	yes		yes	yes	yes	yes
	CP1	MW	0.100	0.849	0.783	1	0.513	0.507	0.002
		Tukey's CI	yes	yes	yes		yes	yes	yes
	CP2	MW	0.329	0.390	0.689	0.513	1	0.988	0.019
		Tukey's CI	yes	yes	yes	yes		yes	yes
	CP3	MW	0.380	0.411	0.761	0.507	0.988	1	0.034
		Tukey's CI	yes	yes	yes	yes	yes		yes
	CP4	MW	0.177	0.001	0.012	0.002	0.019	0.034	1
		Tukey's CI	yes	yes	yes	yes	yes	yes	

At the 11q13 region, the chromatin compaction in the wildtype WP1 differs significantly to that of many of the other cell lines, including the wildtype WP3 and the CdLS lines, CP1, CP2, CP3 and CP4 in the r^2 dataset, but not in the r^2/a dataset. In the r^2/a datasets, CP1 and CP2 have significantly decompacted chromatin compared to all of the wildtype cell lines, except CP2 and WP3 which are not significantly different, as shown by Mann-Whitney tests but not by Tukey's confidence intervals. Other than this, there is little indication of any significant difference between the chromatin compaction of any of the cell lines studied, and no indication of a particular difference between wildtype and CdLS cell lines (Figure 4.19; Table 4.5).

At the 18q22 region it is suggested that the CdLS cell line, CP4, may have significantly decompacted chromatin compared to all other cell lines studied in both the r^2 and r^2/a datasets. However there is no evidence of a particular difference in the chromatin compaction at this region between wildtype and CdLS cell lines (Figure 4.20; Table 4.6).

As with the results in the previous chapter, there is a contrast between the results found when studying the cell lines from Professor Tom Strachan, CN1 and CN2 compared to commercially available W1, and those from the cell lines from Dr Matt Deardorff, the wildtypes WP1, WP2 and WP3, and the CdLS lines CP1, CP2, CP3 and CdL 223P. CN1 and CN2 have decompacted chromatin at regions of high CTCF and cohesin binding compared to W1, whilst CP1, CP2, CP3 and CdL 223P do not have decompacted chromatin at the same regions compared to WP1, WP2 and WP3.

One possible explanation may be the severity of the mutations in these cell lines (Figure 3.1). CN1 has a mutation that skips exon 30 of *NIPBL*, however the protein remains in frame, whilst CN2 has a nonsense mutation in exon 43 of 46 (Tonkin *et al*, 2004). Both of these mutations give a long mRNA which may not be degraded by the nonsense-mediated decay machinery of the cell, resulting in a large, non-functioning protein. In contrast the CdLS cell lines from Philadelphia all have much more severe *NIPBL* mutations giving short mRNAs that would be predicted to be subject to nonsense-mediated decay, therefore giving no peptide: CP4 has a microdeletion that deletes exons 2-17, which includes the start codon; CP2 has an

out-of-frame deletion in exon 3 that results in a premature stop codon; CP1 and CP3 both have mutations in exon 10, 2bp deletion and a point mutation respectively, both of which result in a stop codon. The Newcastle CdLS cell lines could still give a large non-functioning NIPBL protein, whilst the Philadelphia CdLS lines would only give a small non-functioning peptide, if anything. This large non-functioning protein may have a dominant negative effect, whereas the small non-functioning peptide, or no protein, would only give a haploinsufficient effect. Perhaps the decompaction of chromatin I observed in CN1 and CN2 was due to interference of non-functional NIPBL on the activity of the one functioning copy, whilst in the Philadelphia cell lines, the functioning NIPBL was able to act unhindered. Indeed, complete knockout of one copy of *NIPBL* gives a very different phenotype to CdLS (Reeves *et al*, in preparation), suggesting that CdLS is caused by dominant negative effects, not haploinsufficiency. However, all of the cell lines came from patients diagnosed with severe CdLS, suggesting that any differences that can be seen at a cellular level are not apparent at a patient level.

An alternative explanation is that natural variation is responsible for the cellular phenotype observed in CN1 and CN2, but not in CP1, CP2, CP3 and CP4. The wildtype cell lines WP1, WP2 and WP3 were matched controls, obtained from siblings of CdLS patients (though not the same patients as any of the cell lines in this thesis). It is possible that the cellular phenotype I observed in CN1 and CN2 was due to a genetic background effect, and I did not observe it in CP1, CP2, CP3 and CP4 because WP1, WP2 and WP3, having the same genetic background, had the same phenotype.

4.7 Discussion

There is evidence that CdLS is able to affect the compaction of chromatin in different ways at different regions. At a large scale, chromatin in CdLS behaves as if its topology were more homogenous than in wildtype, as tight controls on looping are lost. At a smaller scale, we can see how the complex cohesin interactions at specific regions can result in subtle changes in the compaction of chromatin that differ between CdLS cell lines.

5 RNAi manipulation to replicate the cellular phenotype of CdLS

It was unclear from the experiments I did in the CdLS cell lines, described in Chapter 4, whether the chromatin condensation phenotypes observed in the first set of CdLS cell lines from Newcastle, but not in the second set from the USA, were genuinely due to CdLS or due to some other factor, such as differences in the way cell lines were established in different donor labs. For this reason, I decided to use RNA interference (RNAi) to see if I could replicate these phenotypes by knocking down genes that are involved in CdLS and in cohesin, or that have been implicated in cohesin function.

I chose to knockdown NIPBL, SMC1 and CTCF by RNAi. I considered NIPBL to be the most important choice for knockdown, as this is mutated in 50% of all CdLS cases and was mutated in all of the CdLS cell lines that I studied. NIPBL expression is reduced by 30% in CdLS patients with a *NIPBL* mutation (Liu *et al*, 2009) so I felt that reduction in NIPBL levels by RNAi might replicate the cellular effect of CdLS. I chose to knockdown SMC1 as this is a component of the cohesin complex itself, and is also mutated in some cases of CdLS (Musio *et al*, 2006). I was interested in knocking down CTCF because it is reported to colocalise with cohesin around the genome and is involved in chromosome topology, interacting with cohesin (Section 1.5), therefore the mechanism of disrupted chromatin condensation in CdLS may be through cohesin-CTCF mediated chromatin looping (Wendt *et al*, 2008; Stedman *et al*, 2008; Parelho *et al*, 2008; Zhao *et* al, 2006; Kurukuti *et al*, 2006; Splinter *et al*, 2006; Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010).

5.1 RNAi knockdown of NIPBL , SMC1 and CTCF

I knocked down the three genes in HT1080 human fibroblast cells, as LCLs, which I had used for previous experiments, do not transfect well. HT1080 cells were

derived in 1974 by Rasheed *et al* from a fibrosarcoma from a 35-year old male. These cells have a modal chromosome number of 46, however it is known that 80% display rearranged karyotypes and pseudodiploidy is common. Despite this, the karyotype is stable enough to provide an accurate model of human fibroblasts.

The genes were knocked down for two days using 100nM ON-TARGETplus short interfering RNA (siRNA) SMART-pools ordered from Dharmacon (Section 2.7; Figure 5.1). As well as the three siRNAs targeted to the genes of interest, in each experiment I treated samples of cells with two negative controls, a non-coding SMART-pool from Dharmacon which did not correspond to any endogenous mRNAs, and a mock transfection, using siRNA buffer but no siRNA. It was important to compare coding siRNAs to a non-coding control, because siRNA treatment itself has been shown to affect gene expression in HT1080 cells, particularly genes involved in the immune response, and these changes may also affect compaction (Tagami *et al*, 2005). After 48h, cells were harvested to detect knockdown.

As I was not able to detect NIPBL protein by western blot, perhaps due to its large size (more than 300kDa), I detected knockdown of *NIPBL* mRNA by reverse-transcriptase polymerase chain reaction (RT-PCR; Section 2.4). I was able to demonstrate a qualitative knockdown of *NIPBL* RNA in the NIPBL targeted siRNA treated cells compared to the cells treated with a non-coding control pool siRNA and those mock transfected, by running the PCR products on a 1% agarose gel. qPCR showed that *NIPBL* RNA levels were reduced by more than 90% in the NIPBL-targeted siRNA treated cells compared to the non-coding siRNA treated and the mock transfected cells (Figure 5.2A). In CdLS caused by mutations of *NIPBL*, there is only a loss of 30% of the NIPBL RNA (Liu *et al*, 2009) therefore I have reduced NIPBL expression to below that in CdLS cells, so my model may give a more extreme phenotype than in CdLS.

I was able to detect SMC1 and CTCF proteins by western blot using antibodies from Upstate and Bethyl laboratories respectively, and using a GAPDH antibody from Abcam as a loading control (Section 2.5). SMC1 protein was detected close to the predicted size of 142kDa. Although the predicted mass of CTCF is 83kDa, it is

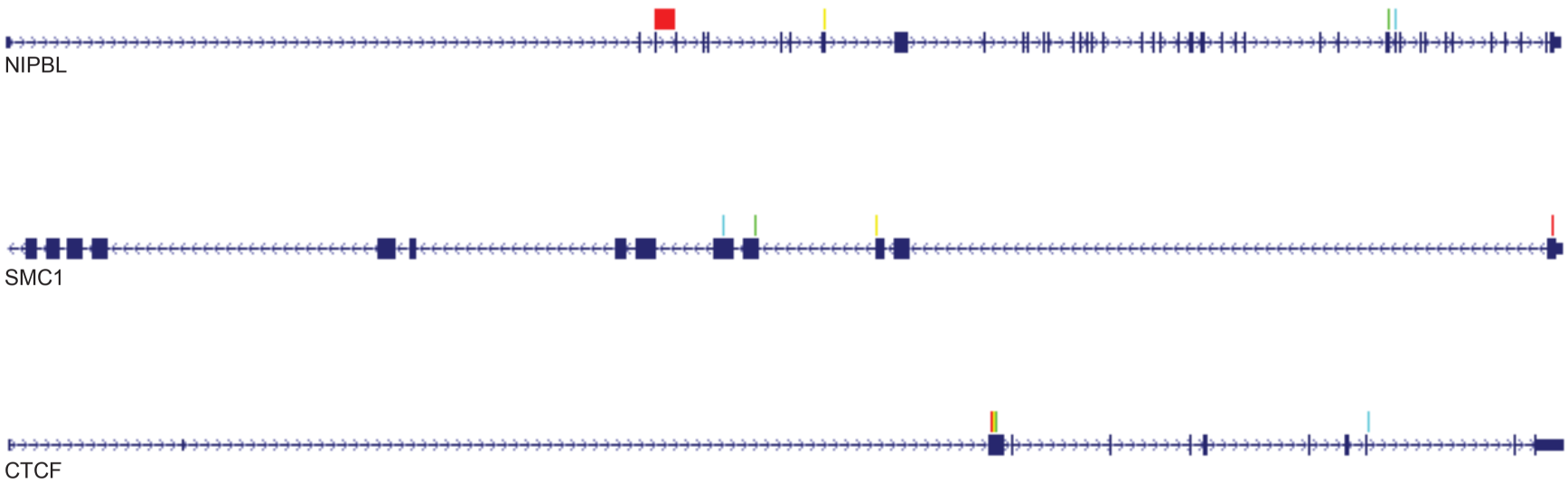
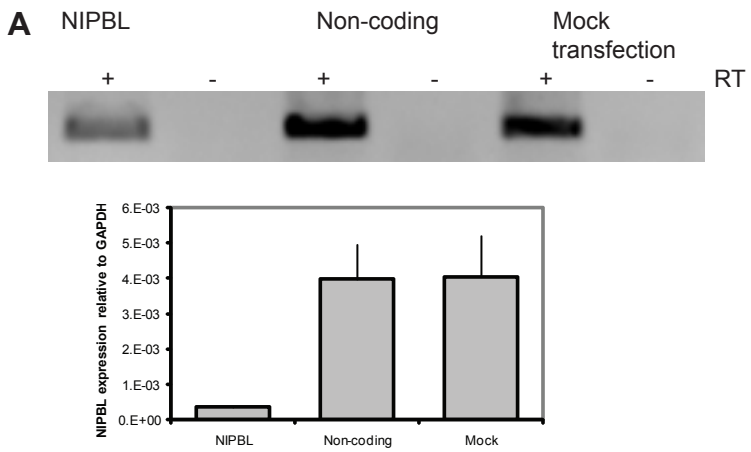
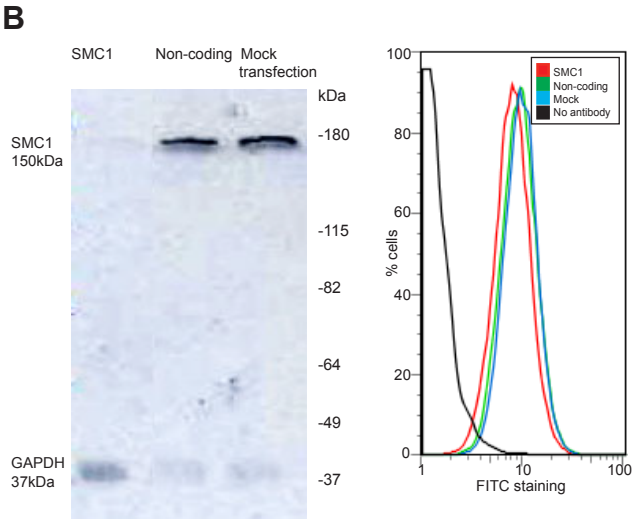
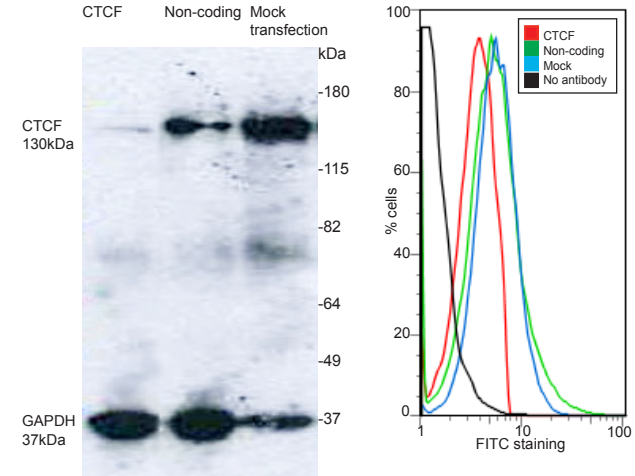


Figure 5.1 - siRNA gene targeting.
Tracks showing where the siRNAs target the genes they knockdown, created using the UCSC genome browser (GRCh37/h19 assembly; Kent et al, 2002; http://genome.ucsc.edu/). The genes are shown in blue with exons in filled boxes. The position of the siRNAs are shown in red, yellow green and cyan above the genes.





	% FITC⁻	mean FITC staining
no antibody	99.74	1.17
SMC1	37.33	8.51
non-coding	24.23	9.9
mock	20.11	10.17



	% FITC⁻	mean FITC staining
no antibody	99.74	1.17
CTCF	92.45	4.1
non-coding	76.39	5.55
mock	70.6	6.09

Figure 5.2 – Knockdown of NIPBL, SMC1 and CTCF by RNAi.
A, levels of NIPBL RNA after RNAi in HT1080 cells, shown by RT-PCR. On the top PCR products are shown on a 1% agarose gel. The graph shows the mean NIPBL mRNA level of three qPCR technical replicates relative to GAPDH, the error bars indicate the standard deviation. **B**, levels of SMC1 protein after RNAi in HT1080 cells, shown by Western Blot and by immunofluorescence FACS. For FACS SMC1 levels were detected in FITC and the FITC levels measured by FACS, shown on a $\log_{10}$ scale; the percentage of FITC negative cells and the mean FITC level were calculated **C**, levels of CTCF protein after RNAi in HT1080 cells, as **B**.

known to migrate at 130kDa in extracts from mammalian cells (Farrar *et al*, 2010). By western blot, I showed a knockdown of both SMC1 (Figure 5.2B) and CTCF (Figure 5.2C) protein levels in the cells treated with the respective siRNAs compared to the non-coding siRNA treated and mock transfected cells. From these westerns, I estimate that I have reduced SMC1 and CTCF proteins levels by 90%. I then carried out FACS analysis of immunofluorescent cells, treating with antibodies against SMC1 and CTCF proteins and staining with FITC, then measured the FITC levels in all cells (Section 2.9.2). There was a general decrease in both SMC1 and CTCF upon knockdown in all cells, not loss of the proteins in some cells and maintained protein levels in other cells, as the histogram of FITC staining was shifted to the left, and not split into two peaks (Figure 5.2B; Figure 5.2C).

5.2 General nuclear phenotype of NIPBL, SMC1 and CTCF knockdown cells

I analysed the RNAi treated cells as I had analysed the CdLS LCLs previously to determine the general nuclear phenotype of the cells. For each experiment I transfected the HT1080 cells with the siRNA of interest (targeted to NIPBL, SMC1 or CTCF) and the non-coding siRNA, and carried out a mock transfection. These cells were harvested after 48h.

5.2.1 Cell cycle analysis of NIPBL, SMC1 and CTCF knockdown cells

I carried out FACS analysis after siRNA mediated knockdown of NIPBL, SMC1 and CTCF, compared to the negative controls, using PI staining as a measure of DNA content (Section 2.9).

There was no significant difference in the cell cycle of the NIPBL knockdown cells compared with either of the two controls (χ^2, $p>0.6$; Figure 5.3). Similarly the cell cycle of the SMC1 knockdown cells was no different to the two controls (χ^2, $p>0.24$; Figure 5.3). In contrast, CTCF knockdown resulted in an increase in the number of cells in G_1, and a corresponding decrease in the number of cells in S-

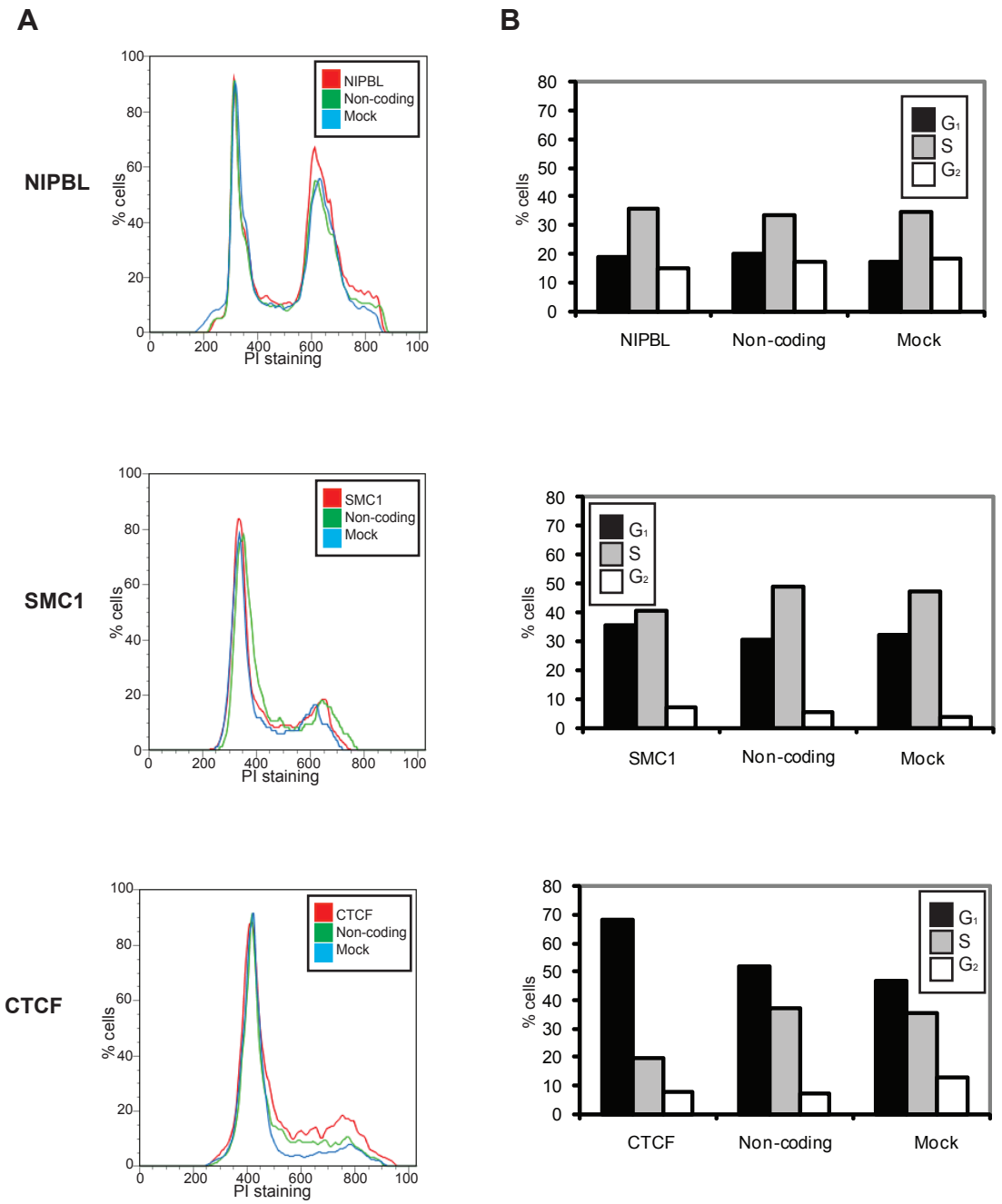


Figure 5.3 – The effect of RNAi knockdown of NIPBL, SMC1 and CTCF on the cell cycle. FACS analysis of three test siRNAs, to NIPBL, CTCF and SMC1, compared to two controls, a non-coding siRNA and a mock transfection. **A**, histograms of DNA content. Cells were excited at 488nm and the fluorescence measured at 562-588nm, the level of fluorescence giving an estimate of DNA content in PI stained cells. **B**, bar charts of the percentage of cells in each cell cycle stage, calculated from the DNA content.

phase compared to both of the controls (Figure 5.3). χ^2 statistical tests suggest that this change was significantly different, at p=0.0001 when comparing CTCF siRNA treatment to both non-coding siRNA treatment and mock transfection. In all three experiments there is evidence of an aneuploid population with a DNA content greater than 2n, which is commonly seen in HT1080 cells, however there was no evidence of increased aneuploidy in any of the three knockdowns compared to controls (Figure 5.3).

It is surprising that SMC1 knockdown does not affect the cell cycle, nor result in increased aneuploidy, given the importance of sister chromatid cohesion in cell cycle progression through mitosis. It may be that residual levels of SMC1 are sufficient for this process.

These data suggest that loss of CTCF may act to trigger a G_1/S checkpoint. It is unlikely that this is due to the putative role of CTCF in positioning cohesin as a similar effect is not observed upon SMC1 knockdown, so it may also be due to other factors, as CTCF is known to interact with a large number of other proteins and to participate in a range of chromatin functions. Conditional knockout of CTCF in T-cells results in cells that are small and do not progress through the cell cycle due to increased expression of cyclin-CDK inhibitors p21 and p27 (Heath *et al*, 2008). Also CTCF knockdown has been previously shown to alter epigenetic marks (Splinter *et al*, 2006; Kurukuti *et al*, 2006), increasing DNA methylation and H3K9me2 and H3K27me2, which may delay S-phase as inactive chromatin tends to replicate later than active chromatin (Hiratani *et al*, 2008). Since similar effects are not seen upon NIPBL or SMC1 knockdown, it is likely that this function of CTCF is independent of cohesin.

5.2.2 Nuclear size of the NIPBL, SMC1 and CTCF knockdown cells

I also analysed the nuclear area after siRNA mediated knockdown to search for evidence of gross changes in nuclear compaction. I fixed the cells in MAA, and visualised the DAPI-stained nuclei using fluorescence microscopy. I analysed the nuclear area of the cells using an IPLab script. It was important to carry out each

experiment simultaneously to control for the effect of humidity on the results. I carried out each experiment twice. The effect of the different treatments on nuclear area in each experiment was compared using Mann-Whitney U-tests.

There is some increase in nuclear area in NIPBL knockdown cells compared to the two controls (Figure 5.4). In the first experiment the mean nuclear area in the NIPBL siRNA treated cells (443.2μm²) was no different than that of the non-coding siRNA treated cells (444.7μm²; $p=0.3218$) but larger than that of the mock transfection (404.4μm²; $p=0.0263$). In the biological replicate the mean nuclear area for the NIPBL siRNA treated cells (485.9μm²) was greater than the two controls, non-coding siRNA (422.1μm²; $p=0.0055$) and mock transfection (357.1μm²; $p<0.0001$).

SMC1 knockdown also resulted in an increase in nuclear area compared to the two negative controls (Figure 5.4). In the first experiment the mean nuclear area of the SMC1 knockdown cells (530.8μm²) was greater than that of the non-coding siRNA treated cells (522.1μm²) and of the mock transfected cells (395.2μm²), whilst in the second experiment the mean nuclear areas were 460.0μm², 385.9μm² and 386.7μm² respectively. In the first experiment there was no significant difference between the nuclear areas after the SMC1 targeted siRNA and non-coding siRNA treatments at $p=0.7985$, but the mock transfected cells had nuclear areas significantly different to both at $p<0.0001$. In contrast, in the second experiment, SMC1 knockdown resulted in a significant difference in nuclear area to the two other treatments at $p<0.0001$.

There is also evidence that CTCF knockdown increases nuclear size (Figure 5.4). The mean nuclear area of the CTCF knockdown cells (303.9μm² in the first experiment and 480.0μm² in the second), was larger than that of the non-coding siRNA treatment (278.1μm² and 478.8μm²), and mock transfection (256.7μm² and 431.8μm²). This is significant at $p<0.0001$ compared to both controls in the first experiment, however in the second experiment CTCF knockdown gives a significant increase compared to the mock transfection at $p<0.0001$, but not compared to the non-coding siRNA at $p=0.9631$. The increase in the nuclear area of CTCF

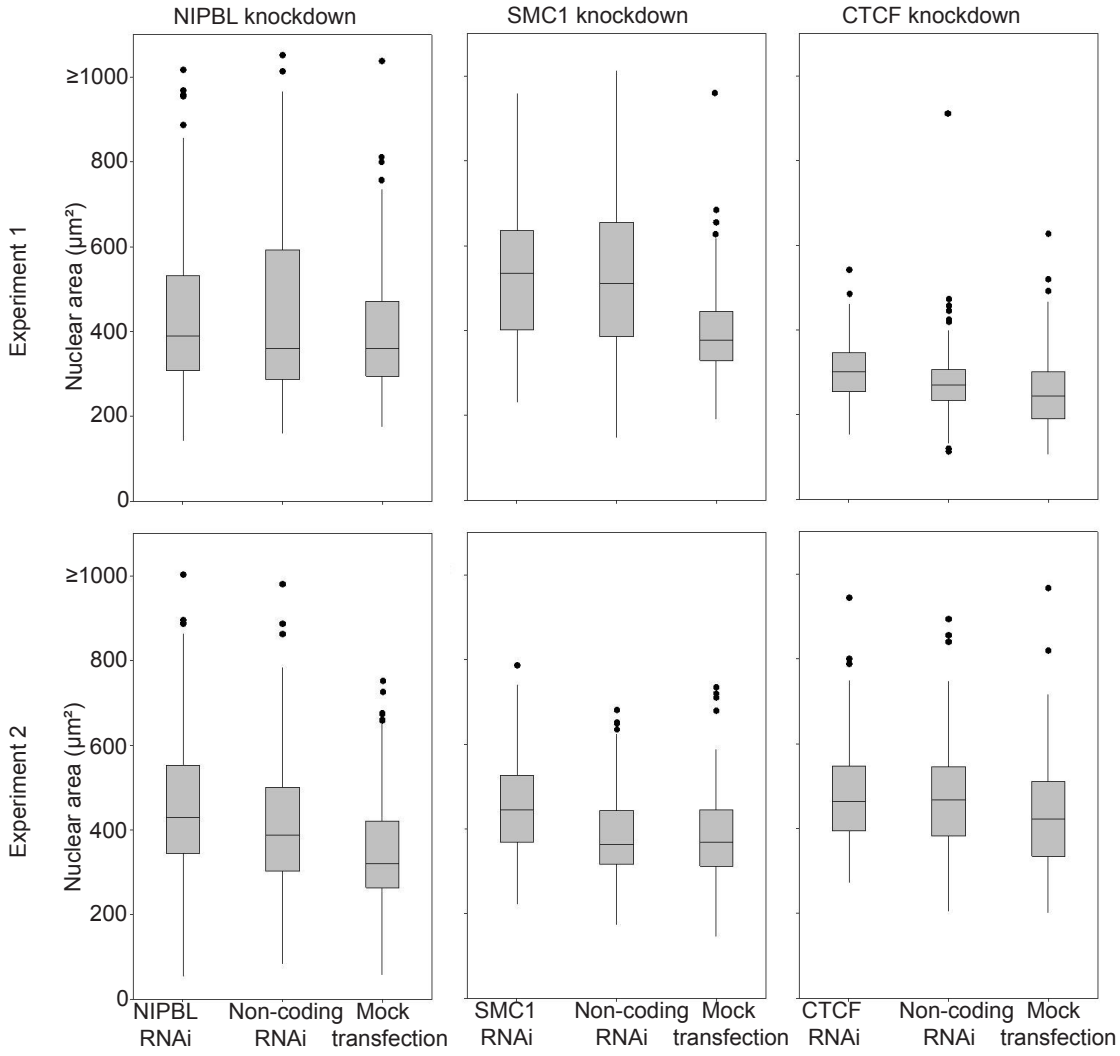


Figure 5.4 – RNAi treatment causes variation in nuclear area as measured by fluorescence microscopy
Boxplots showing the area (µm²) of DAPI staining of MAA fixed cells after RNAi treatment, as boxplots in **Figure 3.2**, n=100. Three test siRNAs, to NIPBL, CTCF and SMC1, were compared to two controls, a non-coding siRNA and a mock transfection in each experiment. Each experiment was carried out twice.

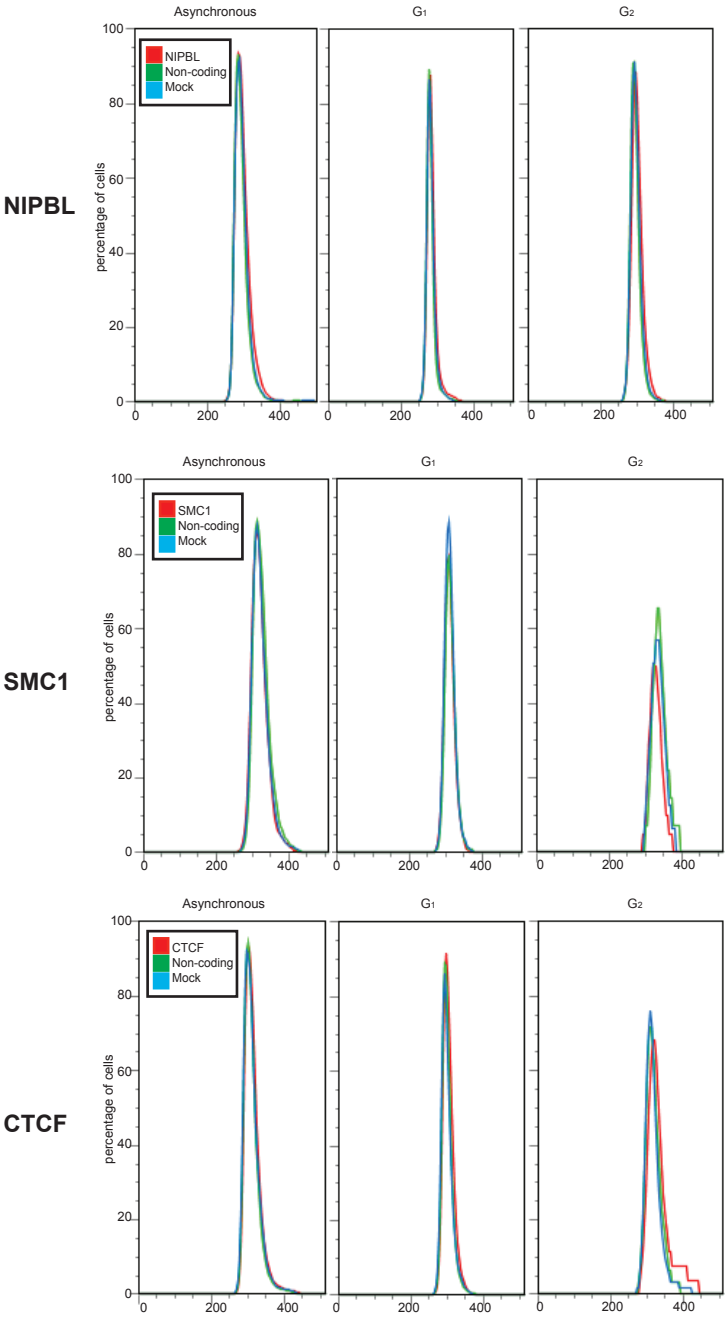
knockdown cells is even more significant when you consider that CTCF knockdowns have a greater proportion of cells in G_1 (Figure 5.3).

There is evidence that knockdown of all three genes results in an increase in nuclear area compared to the mock transfection and sometimes to the non-coding control. However, in all experiments, except the second SMC1 knockdown, the non-coding siRNA also increased nuclear size compared to mock transfection, suggesting that the presence of siRNAs *per se* has some effect on the nuclear area. However, there was an increase in area over and above the increase in the non-coding siRNA treated cells in the NIPBL siRNA, SMC1 siRNA and CTCF siRNA treated cells, suggesting that knockdown of all three results in a gross change in nuclear architecture.

I then analysed nuclear size of siRNA treated cells by width of PI staining in FACS analysis as in Chapter 3. I found no change in PI width after NIPBL and SMC1 siRNA treatments, in neither the asynchronous population nor in G_1 or G_2 only populations. After CTCF knockdown, there appears to be a small increase in the number of larger nuclei in G_2 (Figure 5.5).

5.3 Chromatin compaction at specific chromatin regions in NIPBL, SMC1 and CTCF knockdown cells

To search for more specific changes in chromatin architecture, I decided to study the same regions that I previously analysed in the CdLS cell lines to determine if siRNA knockdown gave the same phenotypes as I observed in some of these cell lines. I chose the CTCF-rich region at 11q13 which showed decompaction in some CdLS lines and the CTCF-poor gene desert at 18q22 which mostly did not differ in compaction between the CdLS and wildtype cell lines, but this region may be even more compact in CdLS than in wildtype (Figure 5.6). I studied these regions by FISH using probe pairs that were ~500kb apart (Section 2.8).



NIPBL

	Mean PI width		
siRNA	Asynchronous	G_1	G_2
NIPBL	295.91	281.16	300.16
Non-coding	293.17	277.8	296.14
Mock	294.69	278.57	297.31

SMC1

	Mean PI width		
siRNA	Asynchronous	G_1	G_2
SMC1	316.88	307.21	328.37
Non-coding	321.88	309.44	339.59
Mock	318.46	307.93	334.55

CTCF

	Mean PI width		
siRNA	Asynchronous	G_1	G_2
CTCF	309.96	300.19	331.56
Non-coding	306.93	297.22	317.63
Mock	307.46	296.22	315.84

Figure 5.5 - RNAi treatment does not affect nuclear size, measured by FACS.
Cells were excited at 488nm and the fluorescence measured at 562-588nm, the width of fluorescence giving an estimate of nuclear size in PI stained cells. Histograms of fluorescence width and mean fluorescence width is shown upon NIPBL, SMC1 and CTCF knockdown, in asynchronous, G_1 and G_2 cells.

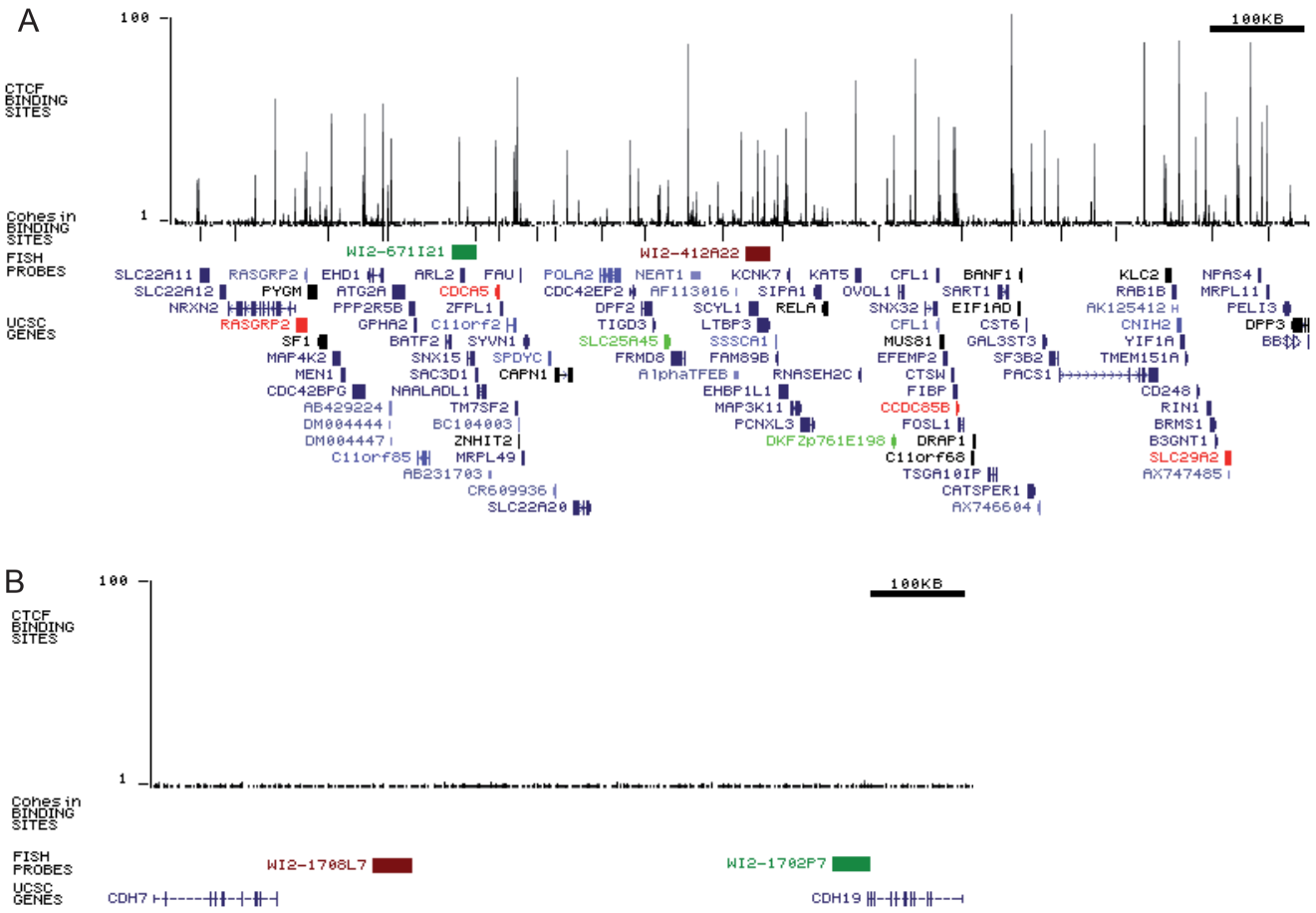


Figure 5.6 - The two regions studied by FISH in the RNAi knockdown cells.
UCSC tracks of 11q13 (**A**; chr11:64,500,000-65,900,000) and 18q22 (**B**; chr18:63,450,000-64,300,000; GRCh37/h19 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). CTCF binding across the region is indicated as a bar chart in black at the top of the track (Barski *et al*, 2007). Cohesin binding across the region is indicated below in black (Liu *et al*, 2009). The positions of the FISH probes used are indicated below, the reference probe is indicated in green and the different distance probes are indicated in red. The genes in the region are in blue, however those that are upregulated in CdLS LCLs compared to wildtype are highlighted in green, and those downregulated in red (Liu *et al*, 2009).

5.3.1 Chromatin compaction of a CTCF-rich chromatin region upon knockdown of NIPBL, SMC1 or CTCF

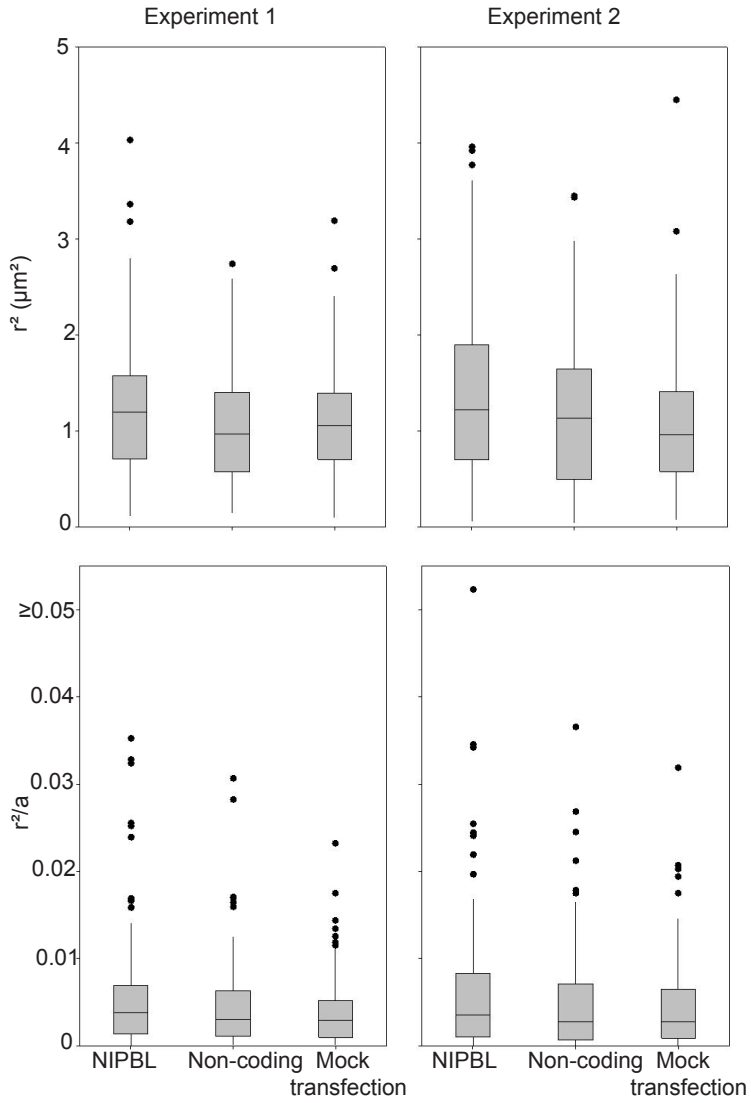
In the previous chapter, I showed that in some CdLS cell lines, the chromatin at 11q13 is significantly decompacted compared to wildtype cells, although this is not the case in all CdLS lines. This region is gene-rich, cohesin-binding rich (Liu *et al*, 2009) and CTCF-binding rich (Barski *et al*, 2007). In addition some of the genes within this region have been shown to be misregulated in CdLS LCLs (Liu *et al*, 2009; Figure 5.6A). None of the genes in this region were shown to be misregulated upon non-coding siRNA treatment in HT1080 (Tagami *et al*, 2008).

Upon NIPBL knockdown there is an increase in r^2 at 11q13 in both biological replicates compared to the two controls, however these increases are not statistically significant by Tukey's confidence intervals, and only compared to the non-coding siRNA in the first experiment by Mann-Whitney U-test, and compared to the mock transfection in the second experiment by Mann-Whitney U-test. However this may be due to a general decompaction of the nucleus as previously observed by this method (Figure 5.4), as in the r^2/a datasets no significant differences between NIPBL and the two controls can be observed; there is a small increase in r^2/a in the NIPBL siRNA treated cells but this is not statistically significant (Figure 5.7).

There is no evidence of any significant change in the compaction of 11q13 upon SMC1 knockdown. In the first replicate, this region appears to be decompacted, as measured by both r^2 and r^2/a, however this is not significant, and there does not appear to be any change in the second replicate (Figure 5.8).

In the first CTCF knockdown experiment, the mock transfected cells have significantly smaller r^2 measurements compared to CTCF and non-coding siRNA treatment, perhaps due to an effect of the treatment itself on the compaction of the nucleus generally, as this effect is erased in the r^2/a datasets. In the second replicate, the non-coding siRNA results in a significant change in r^2, but not in r^2/a compared to both CTCF siRNA and mock transfection, whilst mock transfection gives a significant change in r^2/a, but not in r^2 compared to both CTCF and non-coding siRNA (Figure 5.9).

A

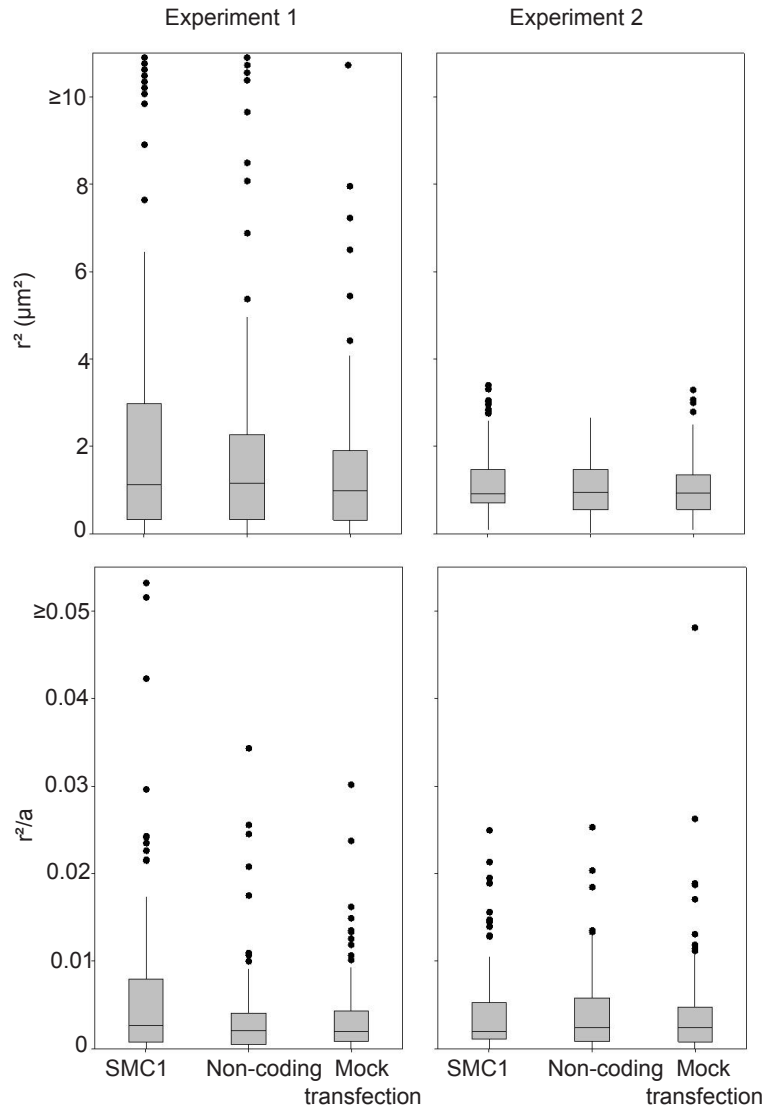


B

Exp1 r²							Exp2 r²					
			NIPBL	Non-coding	Mock					NIPBL	Non-coding	Mock
mean (µm²)			1.22	1.03	1.09		mean (µm²)			1.42	1.18	1.10
median (µm²)			1.20	0.97	1.05		median (µm²)			1.22	1.13	0.96
Compared to	NIPBL	MW	1	0.047	0.210		Compared to	NIPBL	MW	1	0.057	0.006
		Tukey's CI		yes	yes				Tukey's CI		yes	yes
	Non-coding	MW	0.047	1	0.313			Non-coding	MW	0.057	1	0.530
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.210	0.313	1			Mock	MW	0.006	0.530	1
		Tukey's CI	yes	yes					Tukey's CI	yes	yes	
Exp1 r²/a							Exp2 r²/a					
			NIPBL	Non-coding	Mock					NIPBL	Non-coding	Mock
mean			0.0057	0.0047	0.0039		mean			0.0063	0.0048	0.0047
median			0.0038	0.0030	0.0030		median			0.0035	0.0028	0.0028
Compared to	NIPBL	MW	1	0.303	0.063		Compared to	NIPBL	MW	1	0.129	0.176
		Tukey's CI		yes	yes				Tukey's CI		yes	yes
	Non-coding	MW	0.303	1	0.546			Non-coding	MW	0.129	1	0.793
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.063	0.546	1			Mock	MW	0.176	0.793	1
		Tukey's CI	yes	yes					Tukey's CI	yes	yes	

Figure 5.7 – Chromatin compaction at 11q13 after NIPBL siRNA.
A, Boxplots showing the interphase separation squared (r^2) and the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

A

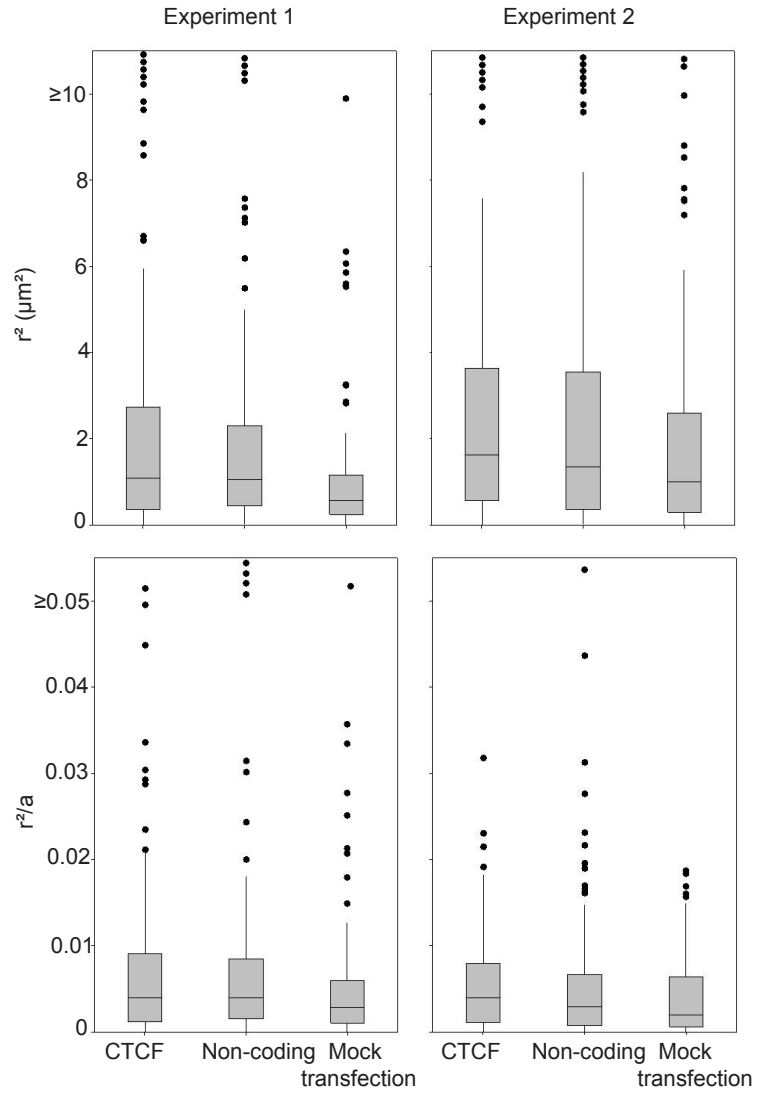


B

Exp1 r²							Exp2 r²					
			SMC1	Non-coding	Mock					SMC1	Non-coding	Mock
mean (µm²)			2.57	2.08	1.48		mean (µm²)			1.16	1.03	1.05
median (µm²)			1.12	1.16	0.99		median (µm²)			0.92	0.95	0.94
Compared to	SMC1	MW	1	0.657	0.216		Compared to	SMC1	MW	1	0.253	0.260
		Tukey's CI		yes	yes				Tukey's CI		yes	yes
	Non-coding	MW	0.657	1	0.375			Non-coding	MW	0.253	1	0.976
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.216	0.375	1			Mock	MW	0.260	0.976	1
		Tukey's CI	yes	yes					Tukey's CI	yes	yes	
Exp1 r²/a							Exp2 r²/a					
			SMC1	Non-coding	Mock					SMC1	Non-coding	Mock
mean			0.0062	0.0035	0.0036		mean			0.0040	0.0041	0.0042
median			0.0027	0.0020	0.0020		median			0.0020	0.0024	0.0024
Compared to	SMC1	MW	1	0.075	0.276		Compared to	SMC1	MW	1	0.817	0.874
		Tukey's CI		yes	yes				Tukey's CI		yes	yes
	Non-coding	MW	0.075	1	0.363			Non-coding	MW	0.817	1	0.686
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.276	0.363	1			Mock	MW	0.874	0.686	1
		Tukey's CI	yes	yes					Tukey's CI	yes	yes	

Figure 5.8 – Chromatin compaction at 11q13 after SMC1 siRNA.
A, Boxplots showing the interphase separation squared (r^2) and the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

A



B

Exp1 r²							Exp2 r²					
			CTCF	Non-coding	Mock					CTCF	Non-coding	Mock
mean (µm²)			2.43	2.14	1.08		mean (µm²)			2.53	2.85	1.93
median (µm²)			1.09	1.06	0.57		median (µm²)			1.63	1.34	1.00
Compared to	CTCF	MW	1	0.675	0.000		Compared to	CTCF	MW	1	0.009	0.096
		Tukey's CI		yes	no				Tukey's CI		yes	yes
	Non-coding	MW	0.675	1	0.000			Non-coding	MW	0.009	1	0.403
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.000	0.000	1			Mock	MW	0.096	0.403	1
		Tukey's CI	no	yes					Tukey's CI	yes	yes	
Exp1 r²/a							**Exp2 r²/a**					
			CTCF	Non-coding	Mock					CTCF	Non-coding	Mock
mean			0.0078	0.0076	0.0052		mean			0.0053	0.0060	0.0041
median			0.0040	0.0040	0.0029		median			0.0040	0.0029	0.0020
Compared to	CTCF	MW	1	0.989	0.104		Compared to	CTCF	MW	1	0.391	0.047
		Tukey's CI		yes	yes				Tukey's CI		yes	yes
	Non-coding	MW	0.989	1	0.055			Non-coding	MW	0.391	1	0.226
		Tukey's CI	yes		yes				Tukey's CI	yes		yes
	Mock	MW	0.104	0.055	1			Mock	MW	0.047	0.226	1
		Tukey's CI	yes	yes					Tukey's CI	yes	yes	

Figure 5.9 – Chromatin compaction at 11q13 after CTCF siRNA.
A, Boxplots showing the interphase separation squared (r^2) and the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

5.3.2 NIPBL and SMC1 knockdown decompact chromatin at a CTCF-poor region, but CTCF knockdown does not

In the previous chapter, I demonstrated that there was no decompaction of the chromatin at 18q22 in CdLS compared to wildtype, indeed in some CdLS cells, the region appears more compact than wildtype, probably due to relative chromatin decompaction at other genomic regions in NIPBL mutants. This region is a gene desert, and is CTCF- and cohesin-binding poor (Barski *et al*, 2007; Liu *et al*, 2009; Figure 5.6B). The flanking genes, CDH7 and CDH19, are not misregulated upon non-coding siRNA transfection in HT0180s (Tagami *et al*, 2008).

In all experiments, r^2 and r^2/a measurements were lower at 18q22 compared to 11q13, as expected from a gene-poor region compared to a ridge.

Surprisingly, given effects at 18q22 in CdLS cells, NIPBL knockdown results in a decompaction of 18q22 compared to the two controls; this effect is significant in both the r^2 and the r^2/a datasets in both experiments, so the decompaction at this region is over and above the gross decompaction of the nucleus (Figure 5.10).

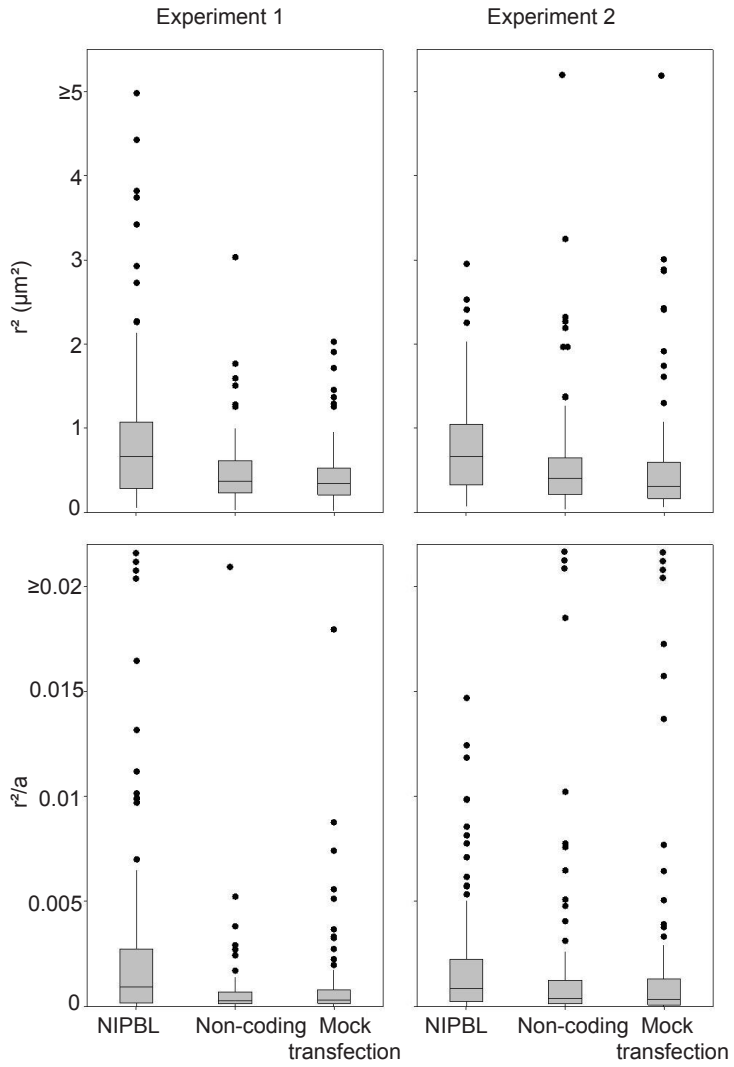
Like NIPBL knockdown, SMC1 knockdown also decompacts chromatin at 18q22. This effect can be seen in the first experiment but is only significant in the r^2 dataset comparing NIPBL knockdown to non-coding siRNA treatment. This effect is significant in both the r^2 and r^2/a datasets in the second experiment (Figure 5.11).

There is no evidence that CTCF knockdown affects the compaction of chromatin at 18q22 compared to the two controls. It appears that this region is compacted in the non-coding control compared to the mock transfection and the CTCF knockdown in both experiments, however this is only significant in the second experiment compared to CTCF knockdown (Figure 5.12). The lack of effect is unsurprising given the absence of CTCF-binding at this region.

5.4 Discussion

In the previous chapter I showed that in some CdLS LCLs the 11q13 region is decompacted compared to wildtype, whereas the 18q22 region remained compact, and indeed the relative level of compaction increased in *NIPBL* mutant cells. In

A



B

Exp1 r²			NIPBL	Non-coding	Mock
mean (µm²)			0.87	0.47	0.45
median (µm²)			0.67	0.38	0.35
Compared to	NIPBL	MW	1	0.000	0.000
		Tukey's CI		no	no
	Non-coding	MW	0.000	1	0.509
		Tukey's CI	no		yes
	Mock	MW	0.000	0.509	1
		Tukey's CI	no	yes	

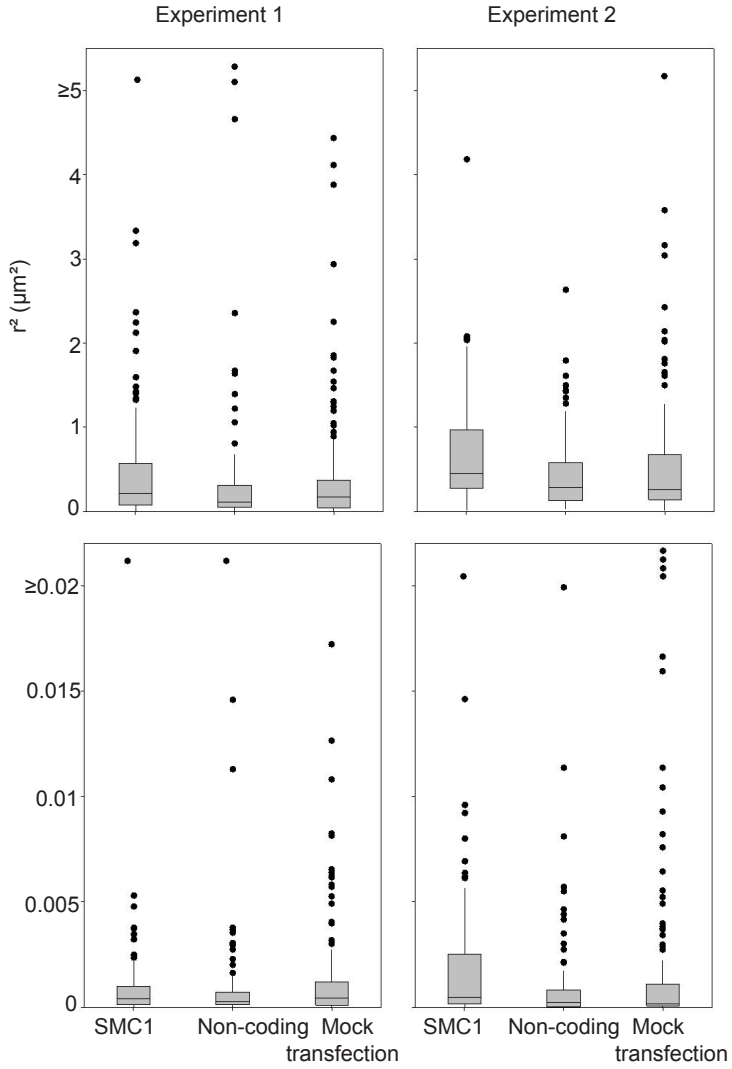
Exp2 r²			NIPBL	Non-coding	Mock
mean (µm²)			0.77	0.61	0.56
median (µm²)			0.67	0.40	0.31
Compared to	NIPBL	MW	1	0.000	0.000
		Tukey's CI		no	no
	Non-coding	MW	0.000	1	0.047
		Tukey's CI	no		yes
	Mock	MW	0.000	0.047	1
		Tukey's CI	no	yes	

Exp1 r²/a			NIPBL	Non-coding	Mock
mean			0.0038	0.0007	0.0009
median			0.0009	0.0003	0.0003
Compared to	NIPBL	MW	1	0.000	0.000
		Tukey's CI		no	no
	Non-coding	MW	0.000	1	0.278
		Tukey's CI	no		yes
	Mock	MW	0.000	0.278	1
		Tukey's CI	no	yes	

Exp2 r²/a			NIPBL	Non-coding	Mock
mean			0.0019	0.0029	0.0039
median			0.0009	0.0004	0.0003
Compared to	NIPBL	MW	1	0.003	0.000
		Tukey's CI		no	no
	Non-coding	MW	0.003	1	0.303
		Tukey's CI	no		yes
	Mock	MW	0.000	0.303	1
		Tukey's CI	no	yes	

Figure 5.10 – Knockdown of NIPBL decompacts chromatin at 18q22.
A, Boxplots showing the interphase separation squared (r²) and the interphase separation squared normalised to nuclear area (r²/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

A

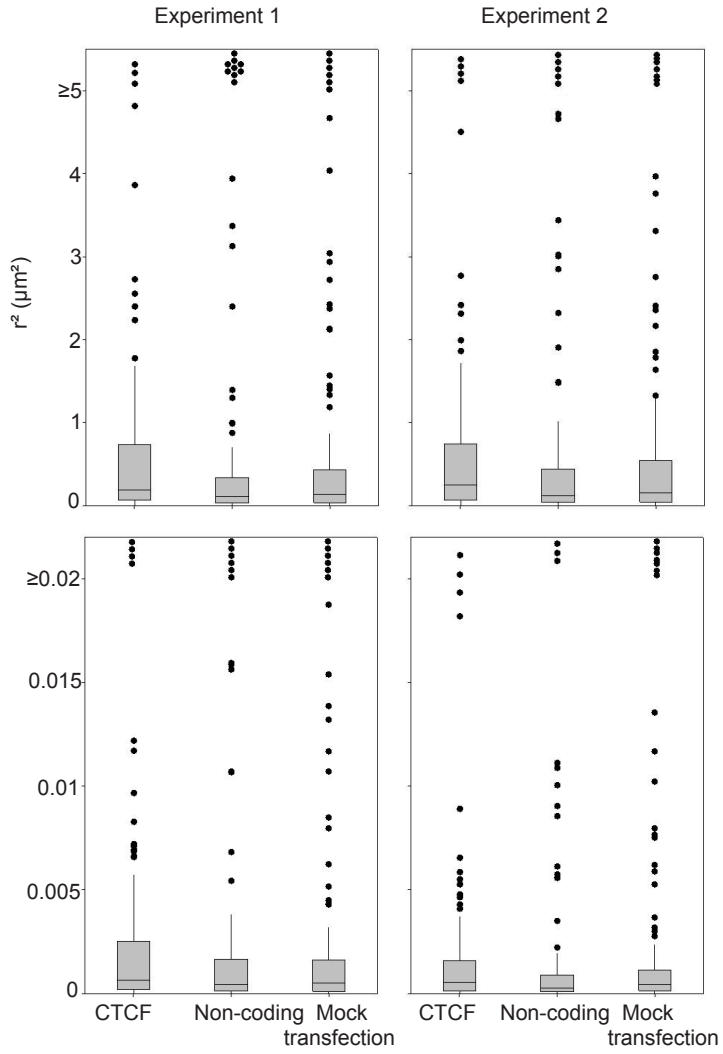


B

Exp1 r²							Exp2 r²					
			SMC1	Non-coding	Mock					SMC1	Non-coding	Mock
mean (µm²)			0.75	0.42	0.46		mean (µm²)			0.67	0.43	0.73
median (µm²)			0.22	0.11	0.17		median (µm²)			0.45	0.29	0.26
Compared to	SMC1	MW	1	0.001	0.085		Compared to	SMC1	MW	1	0.000	0.001
		Tukey's CI		no	yes				Tukey's CI		no	no
	Non-coding	MW	0.001	1	0.258			Non-coding	MW	0.000	1	0.799
		Tukey's CI			yes				Tukey's CI	no		yes
	Mock	MW	0.085	0.258	1			Mock	MW	0.001	0.799	1
		Tukey's CI	yes	yes					Tukey's CI	no	yes	
Exp1 r²/a							Exp2 r²/a					
			SMC1	Non-coding	Mock					SMC1	Non-coding	Mock
mean			0.0015	0.0010	0.0015		mean			0.0019	0.0010	0.0104
median			0.0004	0.0003	0.0005		median			0.0005	0.0003	0.0002
Compared to	SMC1	MW	1	0.187	0.544		Compared to	SMC1	MW	1	0.000	0.007
		Tukey's CI		yes	yes				Tukey's CI		no	no
	Non-coding	MW	0.187	1	0.089			Non-coding	MW	0.000	1	0.603
		Tukey's CI	yes		yes				Tukey's CI	no		yes
	Mock	MW	0.544	0.089	1			Mock	MW	0.007	0.603	1
		Tukey's CI	yes	yes					Tukey's CI	no	yes	

Figure 5.11 – Knockdown of SMC1 decompacts chromatin at 18q22.
A, Boxplots showing the interphase separation squared (r^2) and the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

A



B

Exp1 r²						Exp2 r²					
			CTCF	Non-coding	Mock				CTCF	Non-coding	Mock
mean (µm²)			0.75	1.39	1.07	mean (µm²)			0.91	1.01	1.80
median (µm²)			0.19	0.11	0.14	median (µm²)			0.25	0.12	0.15
Compared to	CTCF	MW	1	0.089	0.303	Compared to	CTCF	MW	1	0.009	0.096
		Tukey's CI		yes	yes			Tukey's CI		yes	yes
	Non-coding	MW	0.089	1	0.625		Non-coding	MW	0.009	1	0.403
		Tukey's CI	yes		yes			Tukey's CI	yes		yes
	Mock	MW	0.303	0.625	1		Mock	MW	0.096	0.403	1
		Tukey's CI	yes	yes				Tukey's CI	yes	yes	
Exp1 r²/a						Exp2 r²/a					
			CTCF	Non-coding	Mock				CTCF	Non-coding	Mock
mean			0.0030	0.0051	0.0046	mean			0.0019	0.0024	0.0051
median			0.0007	0.0004	0.0005	median			0.0005	0.0003	0.0004
Compared to	CTCF	MW	1	0.167	0.288	Compared to	CTCF	MW	1	0.011	0.432
		Tukey's CI		yes	yes			Tukey's CI		yes	yes
	Non-coding	MW	0.167	1	0.890		Non-coding	MW	0.011	1	0.086
		Tukey's CI	yes		yes			Tukey's CI	yes		yes
	Mock	MW	0.288	0.890	1		Mock	MW	0.432	0.086	1
		Tukey's CI	yes	yes				Tukey's CI	yes	yes	

Figure 5.12 – Knockdown of CTCF does not affect chromatin compaction at 18q22.
A, Boxplots showing the interphase separation squared (r^2) and the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2.** This experiment was repeated twice. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and if the Tukey's confidence interval overlap. Significant differences are highlighted in grey.

contrast, I have found that NIPBL and SMC1 knockdown in HT1080 decompacts chromatin at 18q22 to a greater extent than at 11q13.

This apparent paradox can be explained in that the experiments used two different cell types, the B-lymphocyte derived LCLs and the fibroblast derived HT1080s. There will be many differences between the two cell types, including gene expression and chromatin landscape. In the LCLs, the 11q13 region may be tightly regulated by cohesin and NIPBL, and as such is disrupted in CdLS, whilst in the HT1080 cells this region may not be so tightly regulated so is not altered upon SMC1 and NIPBL knockdown, vice versa for the 18q22 region. This suggests that loss of cohesin or NIPBL in CdLS or by RNAi knockdown does decompact chromatin at specific regions, but these regions vary by cell type. Indeed, Liu *et al* (2009) studied gene expression changes in human CdLS LCLs and found genes misregulated at 11q13 but none at 18q22, but there is no data for gene expression in fibroblast-derived CdLS cell lines, which may give the reverse effect. Also, although Liu *et al*, (2009) found no cohesin binding in this region in LCLs, this does not mean that the region is not bound by cohesin in fibroblast-derived cell lines, and it would be interesting to determine this.

CTCF, however, does not decompact chromatin at either of the regions studied. This suggests that the decompaction of chromatin in CdLS is unrelated to the interaction between CTCF and cohesin. Indeed, recent evidence suggests that CTCF may not regulate many tissue specific genes, and cohesin interacts with mediator and tissue specific transcription factors to regulate these genes (Kagey *et al*, 2010; Schmidt *et al*, 2010; Kernohan *et al*, 2010). Given the sparseness of CTCF binding at 18q22, it is unsurprising that CTCF does not affect compaction at this region, but 11q13 compaction is not affected by CTCF either, despite the dense CTCF binding at this region (Barski *et al*, 2007). However this CTCF binding was studied in T-cells, different cell types from the LCLs in which I originally studied CdLS chromatin compaction and from the HT1080 cell lines in which I knocked down CTCF, so may not reflect the chromatin landscape of the cells I have been studying.

It is also possible that the knockdown of NIPBL gives a different phenotype to CdLS due to the level of NIPBL mRNA loss. In CdLS cells, there is only 30% loss

of NIPBL mRNA (Liu *et al*, 2009), however I achieved 90% knockdown of NIPBL mRNA, which may give a different phenotype. Given that complete heterozygous knockout of *NIPBL* gives a different clinical phenotype to CdLS (Reeves *et al*, in preparation), it is entirely possible that a 90% decrease in NIPBL mRNA levels should give a different phenotype to a 30% decrease.

There is no evidence that the genes adjacent to the region studied at 18q22, CDH7 and CDH19 have any particular difference in levels of expression between HT1080 fibrosarcoma-derived cells, or in LCLs (Cheung *et al*, 2005; Lehnhardt *et al*, 2005), however the data available for the two cell lines is not very comparable. The two genes, *CDH7* and *CDH19*, are cadherin genes, involved in cell adhesion (Kools *et al*, 2000) and it seems logical to consider that these genes would be more highly expressed in HT1080s, which grow in an adhesive monolayer, than in LCLs, which grow in a suspension.

6 Bioinformatic analysis of genome-wide expression data in CdLS

The genome-wide effect of *NIPBL* mutation on gene expression has been examined in LCLs derived from human CdLS patients (Liu *et al*, 2009), in two cell lines from a Nipbl knockout mouse model of CdLS (Kawauchi *et al*, 2009), in mouse ES cells with NIPBL and SMC1 knockdown (Kagey *et al*, 2010), and in *Drosophila* cell lines with Nipped-B knockdown (Schaaf *et al*, 2009). In all four studies they found that CdLS or knockdowns that mimic CdLS cause low-level changes (<3-fold) in the expression of a large number of genes rather than a large change of expression of a few genes. In three of the studies they identify candidate genes that may be responsible for the phenotypes observed in CdLS.

I was interested in these studies because I wanted to find regions of the human genome where many genes were misexpressed in CdLS compared to wildtype, perhaps even a gene cluster, as this might be indicative of loss of long-range interaction and changes in chromatin compaction. I was particularly interested in regions that might have genes misexpressed in CdLS in all three species, as these would be regions that were robustly misregulated in a number of biological systems. Such regions would then provide a good focus for further investigation of chromatin structure.

All four studies use microarrays to study gene expression in CdLS in the different species. In humans, gene expression analysis was carried out on 16 LCLs obtained from different CdLS probands, and compared to 17 healthy controls, and 1501 genes were identified that are misregulated with a false discovery rate <0.05 (Liu *et al*, 2009).

A mouse model of CdLS was made by gene-trapping *Nipbl*. This resulted in a 25-30% loss of Nipbl protein in heterozygous animals, similar to the level of NIPBL reduction in CdLS. The phenotype of the mice was similar to that of CdLS, with cranio-facial abnormalities, small size, heart defects and behavioural abnormalities.

Gene expression was analysed in ten *Nipbl*$^{+/-}$ E13.5 brains compared to 11 wildtype from two litters: 978 misexpressed genes were identified in the mutants with a false discovery rate <0.05. In ten *Nipbl*$^{+/-}$ E15.5 MEF lines compared to nine wildtype, from three litters, 81 genes were misexpressed in mutants, with a false discovery rate <0.05 (Kawauchi *et al*, 2009).

A second cellular murine model was created by knocking down Nipbl (75% knockdown) and SMC1 (90% knockdown) in ES cells. It was found that upon Nipbl knockdown 257 genes were downregulated with a $\log_2$ fold change ≤-0.5 and 439 genes upregulated with a $\log_2$ fold change ≥0.5 and upon Smc1 knockdown 818 genes were upregulated and 1310 genes downregulated. These included reduced expression of a number of pluripotency associated markers and increased expression of developmental transcription factors (Kagey *et al*, 2010).

CdLS was modelled in *Drosophila* cell lines by knockdown of Nipped-B by RNAi in the central nervous system (CNS) derived BG3 cell line, and studied after three, four and six days of incubation; 653 genes were found to be misexpressed after three days, 1455 after four days and 1933 after six days siRNA incubation compared to a mock transfection with a $\log_2$ fold change ≤-0.5 or ≥0.5 (Schaaf *et al*, 2009).

6.1 There is no correlation between the genes misregulated in human CdLS cells and the CdLS animal models

In order to find genes that were consistently misregulated in CdLS cells, I asked Dr. Philippe Gautier, in the Bioinformatics department at the MRC Human Genetics Unit to identify orthologous genes that were misregulated in human CdLS and in animal cells which replicate CdLS, from the published data (Liu *et al*, 2009; Kawauchi *et al*, 2009; Kagey *et al*, 2010; Schaaf *et al*, 2009; Section 2.10).

6.1.1 Comparing human CdLS to the mouse model

Only 67 genes were identified that were misregulated in both the mouse gene-trap model (either the brains, or the MEFs; Kawauchi *et al*, 2009) and in the human LCLs (1501; Liu *et al*, 2009). This low number compared to the total number of genes identified in each study (978 in mice and 1501 in humans) suggests that the coincidence of these genes is little more than random. Comparing the mouse ES-cell knockdown data, 142 genes were found in common between human CdLS and Smc1 knockdown, and only 36 in common between human CdLS and Nipbl knockdown (257), with a $\log_2$ ratio of fold change <-0.5 or >0.5 (Liu *et al*, 2009; Kagey *et al*, 2010). A higher proportion of the genes misregulated in each of the knockdowns were also misregulated in human CdLS LCLs, 32.3% and 14.0% of the genes misregulated upon Smc1 and Nipbl knockdown respectively, compared to the mouse gene-trap, however this is still probably random. The level of misexpression relative to wildtype was compared between the two species, and the Spearman's rank correlation coefficient calculated. No correlation was found between the level of increase or decrease in gene expression in the human CdLS LCL data and either the Nipbl knockdown ES cell data, Smc1 knockdown ES cells or the mouse Nipbl gene-trap (ρ=0.390, ρ=0.324, ρ=0.067; Figure 6.1).

Rather than suggesting that the mouse model is not an accurate model of CdLS, this difference between the human and mouse cell lines may represent a difference in cell lines, as the human lymphocyte-derived LCLs came from a very different developmental environment to the fibroblast-derived MEFs and the mouse brain cells, or the pluripotent ES cells, where different genes may be regulated by cohesin. Indeed, in wildtype human LCLs and in mouse ES cells, cohesin is preferentially found near to transcription start sites of genes that are misregulated in CdLS or upon Nipbl knockdown respectively (Liu *et al*, 2009; Kagey *et al*, 2010), suggesting that if cohesin is differentially bound to the genome in different cell lines, then different genes will be misexpressed due to CdLS.

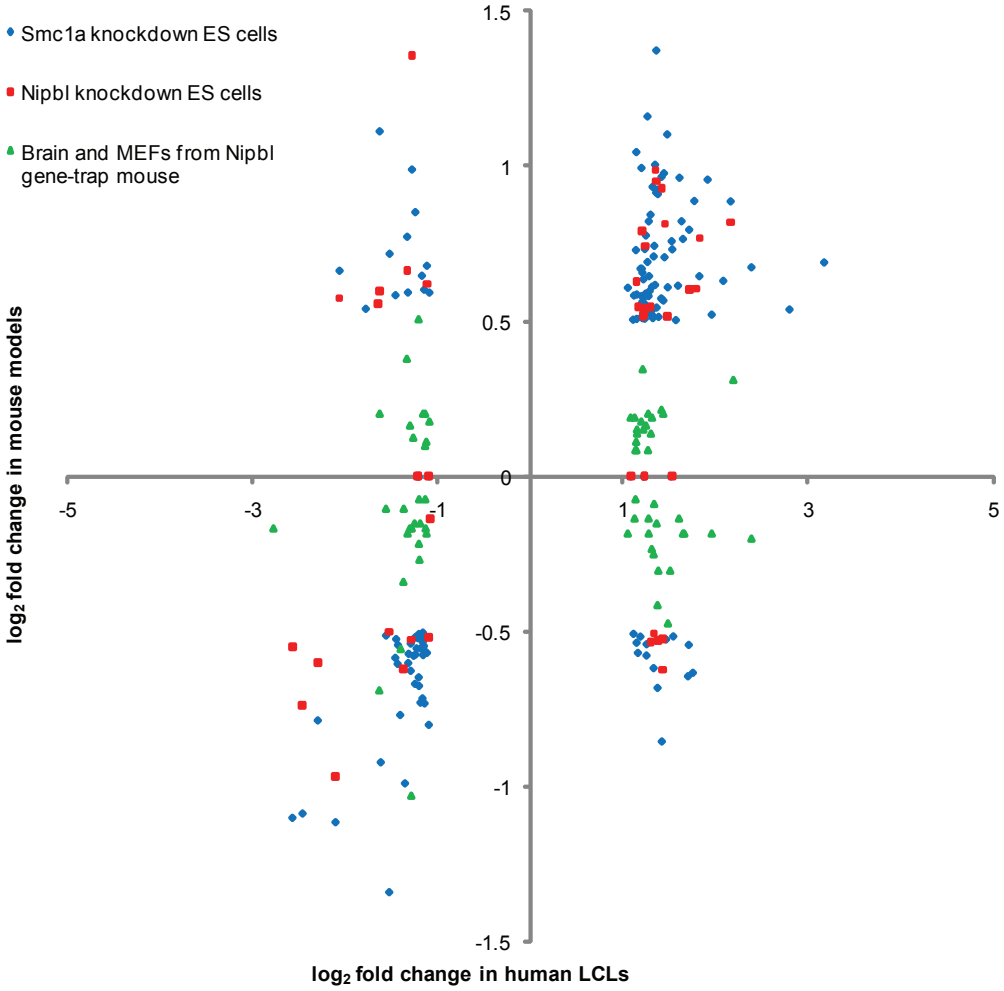


Figure 6.1 - Lack of correlation between the genes misexpressed in CdLS and mouse models of CdLS.

The $\log_2$ fold change in expression of genes that are misexpressed in human CdLS LCLs (Liu *et al*, 2009) is plotted against their fold-change in mouse CdLS models (Kawauchi *et al*, 2009; Kagey *et al*, 2010).

6.1.2 Comparing human CdLS to the *Drosophila* model

There were 320 orthologous genes that were misexpressed in both the human LCLs (Liu *et al*, 2009) and the fly Nipped-B knockdown BG3 cells (Schaaf *et al*, 2009). This is less than 20% of the genes misexpressed in the *Drosophila* cells (1933) and less than 25% of the genes misexpressed in human CdLS cells (1501). Comparison of the level of misexpression in both species by Spearman's rank correlation found that there was no correlation in the level of increase or decrease in gene expression at ρ=0.466 after three days incubation, ρ=0.385 after four days and ρ=0.208 after six days, compared to human LCLs (Figure 6.2). Again we can consider that this is a difference between cell lines, and not necessarily a failure to produce an accurate animal model of CdLS.

6.1.3 There is no correlation in changes in gene expression in different tissues of the CdLS mouse

To determine if the lack of correlation between species I observed was due to differences between cell lines, or differences between species, I asked Dr. Philippe Gautier to compare the two gene-trap mouse samples, the E13.5 brains and the E15.5 MEFs from gene-trap heterozygotes (Kawauchi *et al*, 2009) and the two mouse ES cell knockdown datasets (Kagey *et al*, 2010).

Only 13 genes were found that are misexpressed in both cell types from the gene-trap mice, however there is a strong correlation (ρ=0.850) between the levels of misexpression in both (Figure 6.3A). 24 genes were misexpressed with a $\log_2$ fold change <-0.5 or >0.5 in both the Nipbl and Smc1 knockdown ES cells and there is a strong correlation (ρ=0.721) between the misexpression levels (Figure 6.3B). There is no correlation between the levels of misexpression upon Nipbl and Smc1 knockdown in ES cells and either the Nipbl gene-trap MEFs (ρ=0.067 and ρ=0.038 respectively) or the Nipbl gene-trap brains at (ρ=0.019 and ρ=-0.142 respectively; Figure 6.3C; Figure 6.3D).

The small number of correlated genes suggests that the developmental background of the cells has a huge effect on the genes that may be misexpressed in CdLS. That the misexpression is correlated suggests that genomic context is

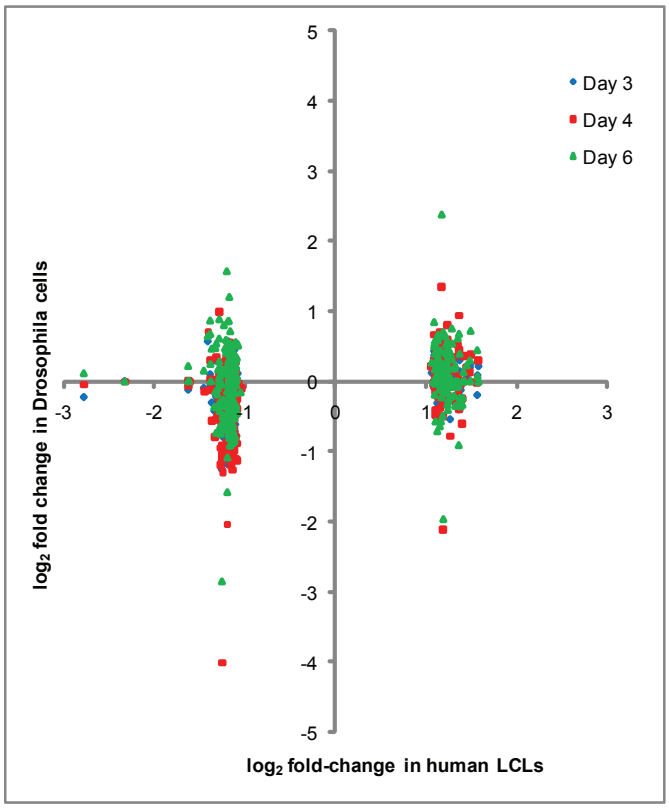


Figure 6.2 - Lack of correlation between the genes misexpressed in CdLS and *Drosophila* CdLS models. The $\log_2$ fold change in expression of genes that are misexpressed in human CdLS LCLs (Liu *et al*, 2009) is plotted against their $\log_2$ fold-change in *Drosophila* CNS cell lines with NIPBL knockdown (Schaaf *et al*, 2009).

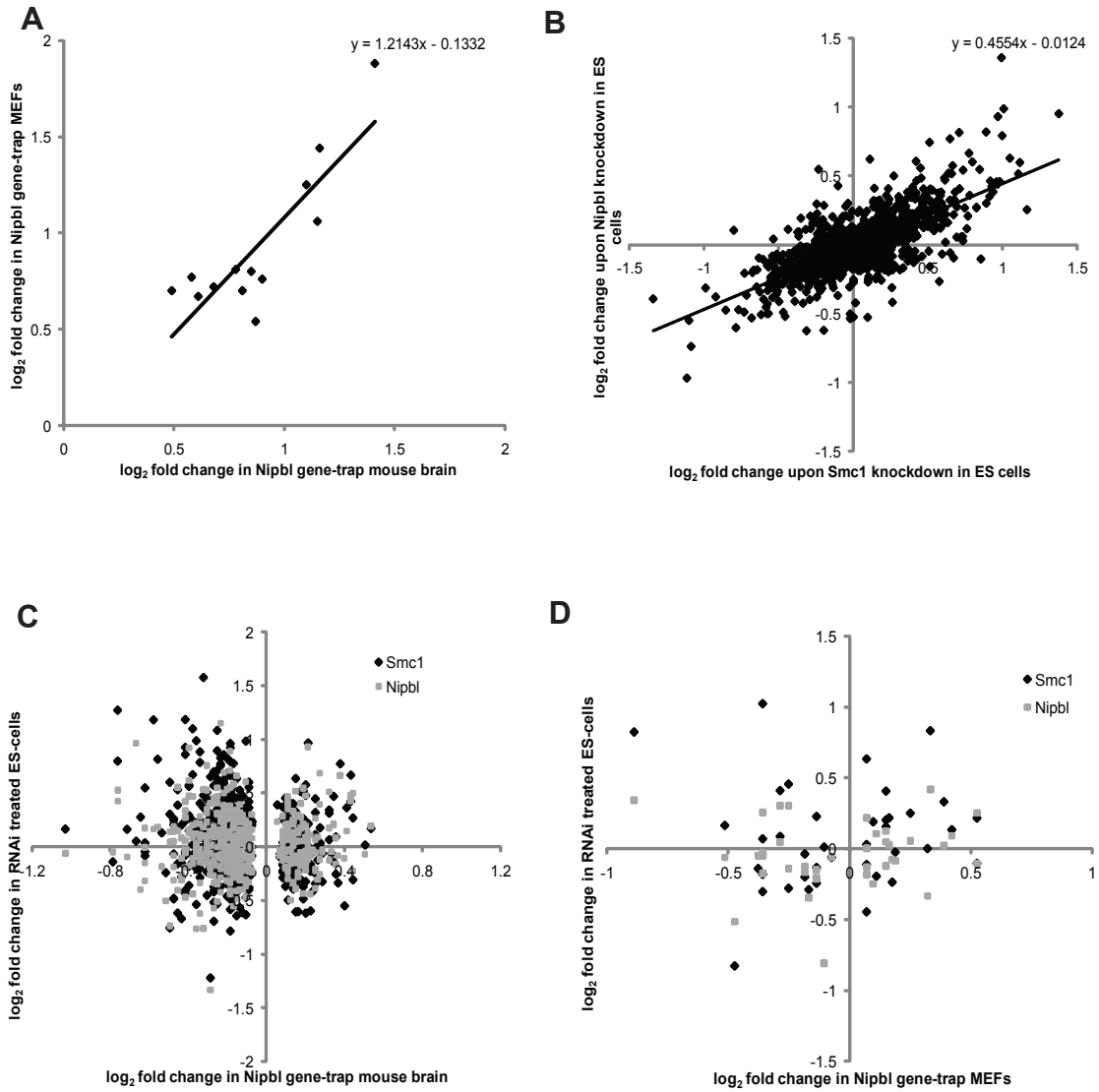


Figure 6.3 - Correlation between misexpression of genes in CdLS mouse models.
The log2 fold change in expression of genes that are misexpressed in different mouse CdLS models. **A**, Comparison of gene expression between MEFs and brain cells from the same Nipbl gene-trap (Kawauchi *et al*, 2009). A line of best fit is shown with the equation of the line, calculated on Microsoft Excel. **B**, Comparison of gene expression between SMC1 and Nipbl knockdown in ES cells (Kagey *et al*, 2010). A line of best fit is shown with the equation of the line, calculated on Microsoft Excel. **C**, Comparison of gene expression between the Nipbl gene-trap mouse brain (Kawauchi *et al*, 2009) and SMC1 and Nipbl ES-cell knockdown (Kagey *et al*, 2010). **D**, Comparison of gene expression between the Nipbl gene-trap MEFs (Kawauchi *et al*, 2009) and SMC1 and Nipbl ES-cell knockdown (Kagey *et al*, 2010).

important in cohesin-mediated gene regulation, however the small number of genes involved suggests that developmental programming may also have a significant effect. These data are similar to those reported by Kagey *et al* (2010) who suggest that cohesin occupies different sites in the genome, and therefore regulates different genes, depending on the tissue. If this is the case, it will have a huge impact on studies on CdLS, cohesin and chromosome topology, which tend to only focus on a single cell line or tissue.

6.2 The protocadherin cluster as a candidate region for chromatin misregulation

Analysis of the three species produced no stand-out candidate region for analysis by FISH. However I decided to look at the protocadherin cluster on human chromosome 5, as many genes in the protocadherin-β cluster were misregulated in the mouse brain CdLS model (Kawauchi *et al*, 2009), and protocadherin-γ C3 was also downregulated in human CdLS LCLs (Liu *et al*, 2009; Figure 6.4A). Protocadherins are expressed in neurons and provide adhesion between neurons at the synapses. The three protocadherin clusters, protocadherin-α, protocadherin-β and protocadherin-γ, consist of a tandem array of differentially spliced exons and a constant region with a set of constant exons. It is thought that the diversity achieved through the different splicing variants leads to synaptic specificity (Yagi, 2008). There is a putative enhancer of protocadherin-α in between protocadherin-α and the protocadherin-β cluster (Ribich *et al*, 2006). A SNP variant within this enhancer is thought to be associated with schizophrenia and bipolar disorder (Pedrosa *et al*, 2008).

In the human genome this region has sparse CTCF and cohesin binding over the protocadherin-α and protocadherin-β clusters, but dense binding of both over the protocadherin-γ cluster in human T-cells and LCLs respectively (Figure 6.4A; Liu *et al*, 2009; Barski *et al*, 2009). In mouse ES cells, there is also more cohesin binding over the protocadherin-γ than over the protocadherin-α and protocadherin-β clusters (Kagey *et al*, 2010; Figure 6.4). Using the UCSC browser I selected two fosmid

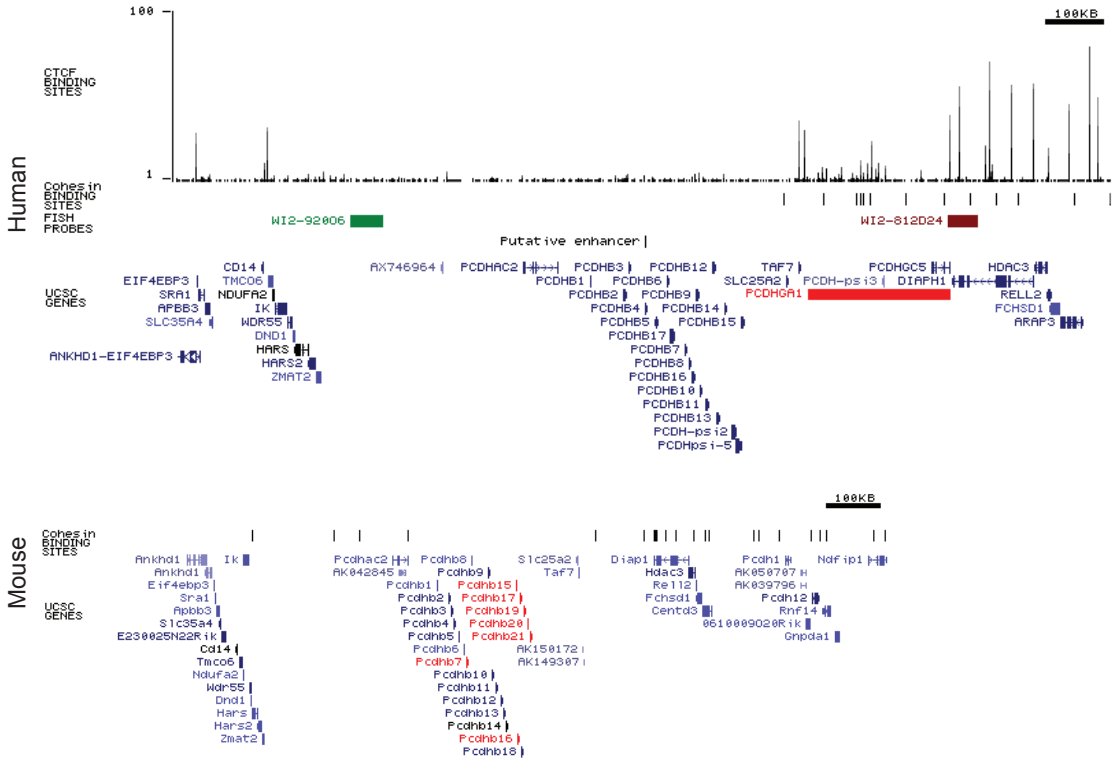


Figure 6.4 - The protocadherin cluster in human and mouse.
UCSC track of the protocadherin region in human (chr5:139,900,000-141,100,000; GRCh37/h19 assembly) and mouse (chr18:36,750,000-38,600,000; NCBI37/mm9 assembly; Kent *et al*, 2002; http://genome.ucsc.edu/). Human, CTCF binding across the region is indicated as a bar chart in black at the top of the track (Barski *et al*, 2007). Cohesin binding across the region is indicated below in black (Liu *et al*, 2009). The positions of the FISH probes used are indicated below, the reference probe is indicated in green and the different distance probes are indicated in red. The genes in the region are in blue, however those that are upregulated in CdLS LCLs compared to wildtype are highlighted in green, and those downregulated in red (Liu *et al*, 2009). Mouse, cohesin binding sites found by ChIP are shown in black (Kagey *et al*, 2010) and genes downregulated in E13.5 mouse brain are highlighted in red (Kawauchi *et al*, 2009).

probes 752kb apart that spanned all three clusters: WI2-920O6, labelled in DIG-11-dUTP and WI2-812D24, labelled in biotin-16-dUTP (Figure 6.4). I measured chromatin compaction by 2D FISH at this region in the CdLS cell lines studied in previous chapters, CN1, CN2, CP1, CP2, CP3 and CdL 223P, compared to the wildtype lines, W1, WP1, WP2 and WP3, as in Chapter 4.

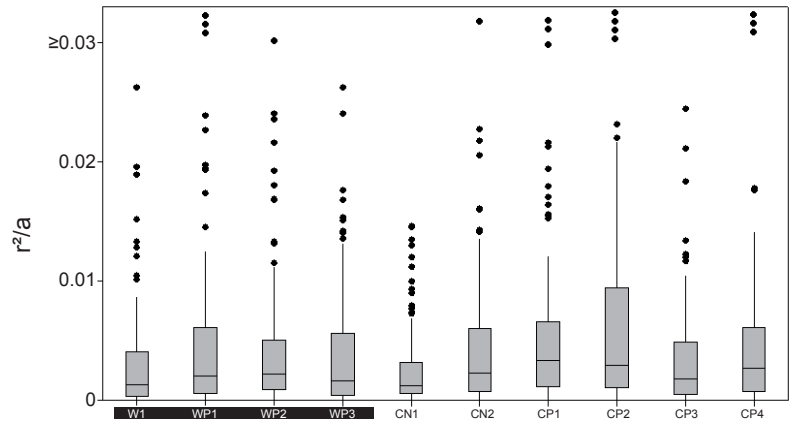
I found that some CdLS cell lines appeared to be compacted at this region compared to wildtype, whilst others appeared to be decompacted (Figure 6.5A). I compared the r^2 and r^2/a datasets pair-wise for each cell line and although I found some pairs of cell lines had significantly different datasets to one another, there was no apparent pattern differentiating CdLS cell lines from wildtype cell lines (Figure 6.4B). From this I conclude that there is no evidence to suggest that the protocadherin region is decompacted in CdLS cell lines compared to wildtype.

Whilst an enhancer has been determined for protocadherin-α, there is not yet evidence of an enhancer of protocadherin-β or protocadherin-γ, which are misregulated in the CdLS mouse model and CdLS respectively. We can speculate that these two clusters have enhancers, and it is the interactions between these enhancers and the promoters of the protocadherin-β and protocadherin-γ genes that are disrupted in CdLS. However, without knowing where these enhancers are, we cannot accurately predict what changes might occur in the compaction of this region. Therefore, it is still likely that the chromatin landscape is altered around the protocadherin cluster in LCLs, but my assay was not able to detect this and a more focussed approach may be necessary.

6.3 Discussion

Authors of papers looking at gene expression in CdLS, be it genome-wide (Liu *et al*, 2009; Kawauchi *et al*, 2009; Schaaf *et al*, 2009) or looking at a specific region (Zhao *et* al, 2006; Kurukuti *et al*, 2006; Splinter *et al*, 2006; Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010) are keen to identify specific genes which may be linked to one or more of the phenotypes of CdLS. However, in the small number of cell lines in which genome-wide gene expression has been studied, there are a large number of genes subtly but

A



B

r²

			W1	WP1	WP2	WP3	CN1	CN2	CP1	CP2	CP3	CP4
mean (µm²)			0.73	0.99	1.19	0.76	0.57	1.00	1.33	1.53	0.70	0.91
median (µm²)			0.28	0.47	0.46	0.39	0.30	0.49	0.70	0.62	0.37	0.63
Compared to	W1	MW	1	0.0611	0.0035	0.5455	0.8184	0.0298	0.0002	0.0008	0.4132	0.0046
		Tukey's CI		yes	yes	yes	yes	yes	no	yes	yes	yes
	WP1	MW	0.0611	1	0.4111	0.1871	0.0497	0.8839	0.0563	0.46	0.2162	0.4315
		Tukey's CI	yes		yes	yes	yes	yes	yes	yes	yes	yes
	WP2	MW	0.0035	0.4111	1	0.0346	0.0032	0.4397	0.2618	0.6013	0.0278	0.8656
		Tukey's CI	yes	yes		yes	yes	yes	yes	yes	yes	yes
	WP3	MW	0.5455	0.1871	0.0346	1	0.576	0.1468	0.0021	0.0058	0.8758	0.0357
		Tukey's CI	yes	yes	yes		yes	yes	no	yes	yes	yes
	CN1	MW	0.8184	0.0497	0.0032	0.576	1	0.0414	0.0001	0.0008	0.5116	0.0047
		Tukey's CI	yes	yes	yes	yes		yes	no	no	yes	yes
	CN2	MW	0.0298	0.8839	0.4397	0.1468	0.0414	1	0.0794	0.1749	0.194	0.396
		Tukey's CI	yes	yes	yes	yes	yes		yes	yes	yes	yes
	CP1	MW	0.0002	0.0563	0.2618	0.0021	0.0001	0.0794	1	0.8075	0.0025	0.3656
		Tukey's CI	no	yes	yes	no	no	yes		yes	yes	yes
	CP2	MW	0.0008	0.46	0.6013	0.0058	0.0008	0.1749	0.8075	1	0.0066	0.4132
		Tukey's CI	yes	yes	yes	yes	yes	yes	yes		yes	yes
	CP3	MW	0.4132	0.2162	0.0278	0.8758	0.5116	0.194	0.0025	0.0066	1	0.0406
		Tukey's CI	yes	yes	yes	yes	yes	yes	yes	yes		yes
	CP4	MW	0.0046	0.4315	0.8656	0.0357	0.0047	0.396	0.3656	0.4132	0.0406	1
		Tukey's CI	yes	yes	yes	yes	yes	yes	yes	yes	yes	

r²/a

			W1	WP1	WP2	WP3	CN1	CN2	CP1	CP2	CP3	CP4
mean			0.0030	0.0049	0.0043	0.0040	0.0027	0.0046	0.0053	0.0079	0.0033	0.0046
median			0.0013	0.0022	0.0022	0.0016	0.0012	0.0022	0.00336080	0.0030	0.0018	0.0027
Compared to	W1	MW	1	0.0514	0.0078	0.1406	0.5425	0.0021	0	0	0.261	0.0011
		Tukey's CI		yes	yes	yes	yes	yes	no	no	yes	no
	WP1	MW	0.0514	1	0.5814	0.6385	0.1323	0.2964	0.0274	0.089	0.3359	0.2425
		Tukey's CI	yes		yes	yes	yes	yes	yes	yes	yes	yes
	WP2	MW	0.0078	0.5814	1	0.2925	0.0207	0.6334	0.0729	0.0747	0.1462	0.3796
		Tukey's CI	yes	yes		yes	yes	yes	yes	yes	yes	yes
	WP3	MW	0.1406	0.6385	0.2925	1	0.3289	0.105	0.0092	0.0063	0.71	0.9043
		Tukey's CI	yes	yes	yes		yes	yes	yes	yes	yes	yes
	CN1	MW	0.5425	0.1323	0.0207	0.3289	1	0.0055	0.0001	0.0001	0.5487	0.003
		Tukey's CI	yes	yes	yes	yes		yes	no	no	yes	yes
	CN2	MW	0.0021	0.2964	0.6334	0.105	0.0055	1	0.2077	0.2293	0.0418	0.7782
		Tukey's CI	yes	yes	yes	yes	yes		yes	yes	yes	yes
	CP1	MW	0	0.0274	0.0729	0.0092	0.0001	0.2077	1	0.9016	0.0019	0.3276
		Tukey's CI	no	yes	yes	no	yes	yes		yes	yes	yes
	CP2	MW	0	0.089	0.0747	0.0063	0.0001	0.2293	0.9016	1	0.0015	0.2555
		Tukey's CI	no	yes	yes	no	yes	yes	yes		yes	yes
	CP3	MW	0.261	0.3359	0.1462	0.71	0.5487	0.0418	0.0019	0.0015	1	0.0312
		Tukey's CI	yes	yes	yes	yes	yes	yes	yes	yes		yes
	CP4	MW	0.0011	0.2425	0.3796	0.9043	0.003	0.7782	0.3276	0.2555	0.0312	1
		Tukey's CI	no	yes	yes	yes	yes	yes	yes	yes	yes	

Figure 6.5 - Chromatin compaction does not change at the protocadherin cluster in CdLS. **A**, Boxplot showing the interphase separation squared normalised to nuclear area (r^2/a), as **Figure 3.2**. Wildtype cell lines are highlighted in black. **B**, tables of statistics, showing the mean and median of each dataset, and the results of pair-wise Mann-Whitney U-tests and Tukey's confidence interval overlaps. Wildtype cell lines are highlighted in black and significant differences are highlighted in grey.

significantly misregulated and I have shown that there is a huge variation in which genes are misregulated and to what extent between the different species and cell lines or tissues studied. I consider it likely that almost all genes in the genome may be misregulated in CdLS, in one tissue or another. Given this variation, it can only be of interest if one particular gene is highly misregulated in CdLS in the tissue in which it is normally active, for example the loss of expression of pluripotency-associated genes in ES cells upon Nipbl knockdown (Kagey *et al*, 2010), or loss of expression of protocadherins in the developing brain (Kawauchi *et al*, 2009). Otherwise, we can consider that CdLS is caused by small alterations in the expression of most genes, varying between tissue types, giving a compound effect.

7 Discussion

In this thesis I have discussed the effects of CdLS on chromatin by studying compaction at various regions in CdLS LCLs and by RNAi of proteins involved in CdLS, the cohesin complex and chromatin architecture. I found that at the scale of chromosome territories, chromatin organisation tends towards a homogenisation in CdLS, with smaller than average CTs increasing in size and larger than average CTs decreasing (Section 4.1), and regions usually internal to the CT moving closer to the periphery and external regions moving into the CT (Section 4.2). On a smaller scale, changes in chromatin compaction in CdLS will depend upon the particular interactions of cohesin at that region (Section 4.3); however the extent of these changes may depend upon the CdLS mutation, with more severe protein truncations tending to give a weaker cellular phenotype than mild truncations (Section 4.6). Though not seen in all cell lines to a statistically significant level, gene-rich regions with dense cohesin and CTCF binding tend to decompact in CdLS, whereas gene-poor regions do not. (Section 4.3; Section 4.6). The phenotypes of CdLS vary between tissues: there is misexpression of different genes in different tissues (Section 6.1.3) and the effects on chromatin compaction differ between CdLS LCLs and RNAi knockdown HT1080s (Section 5.3).

7.1 Cohesin regulates gene expression independently of CTCF

CTCF is not the only partner of cohesin-mediated DNA interactions as CTCF knockdown does not give the same chromatin phenotype as NIPBL and SMC1 knockdown (Section 5.3).

It has been found that CTCF interacts with cohesin at 79% of genomic loci (Wendt *et al*, 2008; Parelho *et al*, 2008; Stedman *et al*, 2008). These interactions result in the looping of chromatin in ACHs (active chromatin hubs) detected by 3C, mediated by CTCF and tethered by cohesin, that can regulate gene expression

(Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010).

However I have found that chromatin at 18q22 is decompacted upon knockdown of NIPBL and SMC1 in fibroblast derived cells, but not upon CTCF knockdown (Section 5.3). Thus I conclude that changes in chromatin structure mediated by cohesin are independent of CTCF. This can be added to growing body of data that suggests that cohesin is able to regulate gene expression through chromatin topology by interacting with proteins other than CTCF such as Mediator, and some tissue specific proteins, such as ATRX, ER, HNF4A and CEBPA (Kagey *et al*, 2010; Kernohan *et al*, 2010; Schmidt *et al*, 2010). It has also been shown that NIPBL only associates with cohesin-binding sites where CTCF does not also interact, suggesting that CTCF-mediated interactions may not be altered in CdLS (Kagey *et al*, 2010). It is likely that there are a number of other tissue specific transcription factors that are able to bring about intrachromosome interactions that can be tethered by cohesin to regulate gene expression (Figure 7.1).

To identify other proteins that can mediate interactions with cohesin it may be necessary to carry out immuno-precipitations with the cohesin proteins in a variety of tissues, and to then determine co-occupation of DNA loci by these proteins by ChIP, and determine any intrachromosome interactions between these loci by 3C or 5C.

These data lead me to ask: how important is CTCF in regulating chromosome topology with cohesin? At a number of regions it has been demonstrated that CTCF and cohesin cooperate to form chromosome loops that regulate gene expression (Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010) so we cannot exclude CTCF-cohesin interaction as a method of gene regulation. However CTCF association and NIPBL association at cohesin binding sites are mutually exclusive (Kagey *et al*, 2010), suggesting that while CTCF can regulate intra-chromosome interactions with cohesin, it is not these interactions that are perturbed in CdLS. Indeed, cohesin co-localises with NIPBL and Mediator together, although this relationship has not been shown with the tissue specific transcription factors, so it is the intra-chromosome interactions mediated by this trio that are disrupted in CdLS.

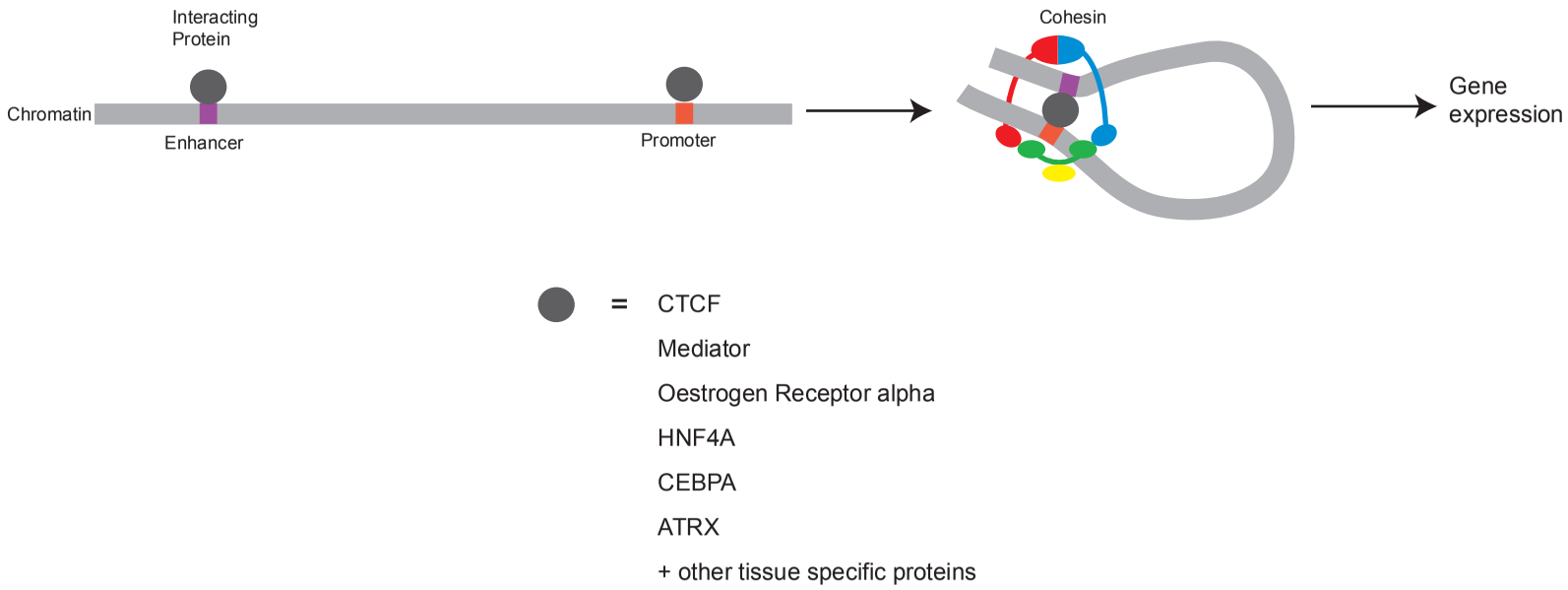


Figure 7.1 - Cohesin interacts with a variety of proteins to regulate gene expression.
Cartoon showing how tissue specific proteins may interact with promoters and enhancers, bringing these into contact via looping of the chromatin. These loops may then be tethered by cohesin, leading to expression of tissue specific genes. Proteins that have been identified in this role are indicated (Kagey *et al*, 2010; Schmidt *et al*, 2010; Nativio *et al*, 2009; Kernohan *et al*, 2010).

7.2 NIPBL mutations have dominant negative effects

My data suggest that there may be a stronger effect on chromatin compaction in cell lines where the *NIPBL* mutation has a smaller effect on the resulting protein, i.e. C-terminal mutations may have a larger cellular phenotypic effect than N-terminal mutations (Chapter 4). This suggests a negative correlation between severity of the mutation and the severity of the cellular phenotype. Clinical studies found that patients who present with mild CdLS phenotypes are more likely to have mild *NIPBL* mutations, such as missense mutations, and patients with severe CdLS often have truncating *NIPBL* mutations, such as frameshift and nonsense mutations, however this correlation is not absolute (Gillis *et al*, 2004; Bhuiyan *et al*, 2006; Selicorni *et al*, 2007). These studies, however, did not look at the severity of the truncations within this category, and all of the CdLS LCLs I studied had a truncating mutation (Figure 3.1). It is possible that missense mutations give a mild phenotype because they give an almost functional NIPBL protein which is able to load cohesin onto chromatin at near-wildtype efficiency. C-terminal truncating mutations should produce much larger, though probably non-functional, NIPBL proteins which could interfere with the activity of wildtype protein, either preventing loading of cohesin onto chromatin or loading cohesin onto different loci; it is conceivable that a larger non-functional protein could have a stronger negative effect than a smaller non-functional protein (Figure 7.2).

NIPBL is a large protein of unknown structure, however it contains two domains of interest: a PxVxL domain which can interact with chromoshadow domains (Lechner *et al*, 2005), and HEAT repeats which act as a protein binding scaffold (Neuwald and Hirano, 2000; Figure 3.1). The mutated proteins produced in CN1 and CN2 cell lines, which both showed a strong effect on chromatin compaction, are both large proteins that retain both the PxVxL domains and the HEAT repeats, although CN1 has an in-frame skipped exon in between the second and third HEAT repeats. CP1, CP2 and CP3, however, all produce short NIPBL peptides which lack both the PxVxL domain and the HEAT repeats, whilst CP4 produces no peptide at all, and these four cell lines had a smaller difference in chromatin compaction to wildtype (Figure 3.1). If these two domains are present they may allow the mutant NIPBL to interact with some protein partners, but the mutations may prevent other interactions

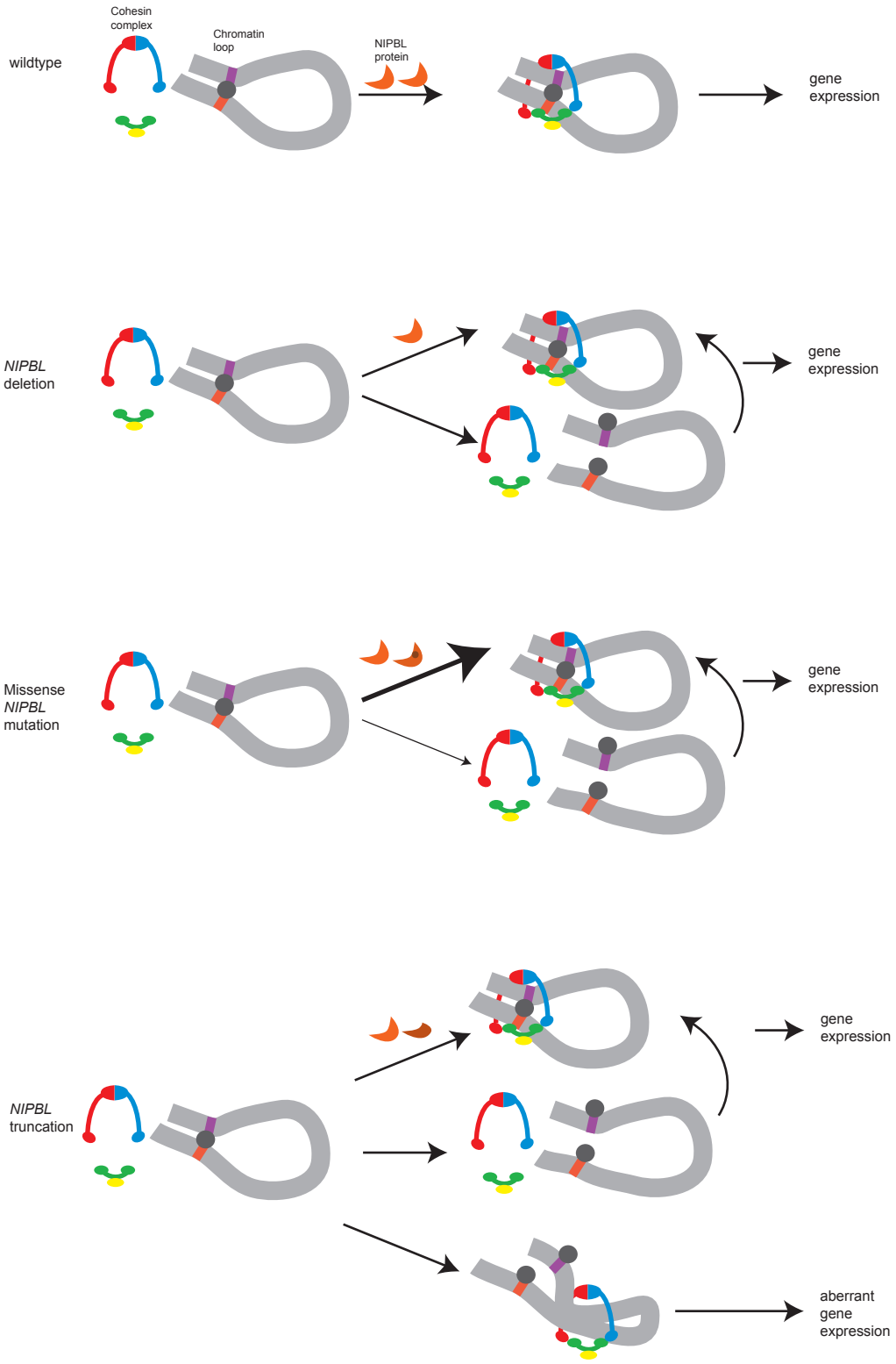


Figure 7.2 - How *NIPBL* mutations might have a dominant negative effect.
Cartoon illustrating how mutated NIPBL proteins may affect tethering of chromatin in different ways.

or enzymatic activity, thus mutant NIPBL may be able to block access of wildtype NIPBL to partner proteins, having a dominant negative effect. Smaller truncated proteins, without the PxVxL domain and the HEAT repeats, will not be able to interact with so many proteins, so the dominant negative effect may be reduced.

These data suggest that *NIPBL* mutations might have a dominant negative effect, as opposed to a haploinsufficient effect as previously suggested (Krantz *et al*, 2004; Tonkin *et al*, 2004). A dominant negative effect of *NIPBL* mutation is also suggested by the non-CdLS phenotype recently observed in a patient with a complete *NIPBL* deletion (Reese *et al*, in preparation); since this patient possessed none of the facial phenotypes observed in CdLS, it is likely that these phenotypes are caused by a dominant negative effect of mutant NIPBL, and not a haploinsufficient effect.

In the case of the knockdown I carried out with NIPBL, this would not give a dominant negative effect, as only the quantity of protein was reduced. However I achieved over 90% knockdown of NIPBL mRNA upon RNAi treatment, compared to the only ~30% loss observed in CdLS cell lines (Liu *et al*, 2009). This suggests that NIPBL mutation may have both dominant negative and haploinsufficient effects, that loss of a high proportion of the endogenous protein results in decompaction of chromatin, as does interference of non-functional protein with functional protein, but that loss of only a small proportion of the endogenous protein is not sufficient to produce a similar effect.

It would be interesting if the interacting partners of mutated NIPBL proteins could be determined by immuno-precipitation with tagged mutant and wildtype proteins, to discover which of the different mutants are able to interact with different NIPBL protein partners. After this, the dominant negative effect on intrachromosome interactions could be assayed by 5C, comparing interactions at a region known to be disrupted in CdLS. Chromosome interactions could be compared between different CdLS cell lines, with a variety of *NIPBL* mutations, including missense, truncating mutations giving a large protein, truncating mutations giving a small protein and complete deletion. Quantitative analysis could identify the frequency of interactions in the different mutants compared to wildtype and determine which kinds of mutation have the strongest effect on interactions.

Searching the literature, I have been unable to identify another dominant negative genetic disorder where a large non-functional protein would have a stronger effect than a small non-functional protein. However, this does not mean this idea is not plausible.

7.3 FISH as an assay to study CdLS

Throughout my PhD I have used the 2D FISH assay to study chromatin compaction in CdLS. Although I have found some effects of CdLS on chromatin compaction, I have not been able to identify the specific interactions that take place at the regions studied and therefore found the effects at some of the regions studied difficult to interpret.

I think that to study chromatin in CdLS further it would be important to study it at a molecular level, and the most useful assay may be by the C methods (Figure 1.7). I think that 5C is the best method for studying interactions of known genomic regions as it is a higher throughput method than 3C, which has already been used in a number of studies identifying CTCF or Mediator and cohesin mediated interactions (Nativio *et al*, 2006; Degner *et al*, 2009; Hadjur *et al*, 2009; Mishiro *et al*, 2009; Hou *et al*, 2010; Kagey *et al*, 2010). Hi-C, however, is a developing technology that is able to detect genome wide interactions, but has not yet been used to study interactions at such high resolution as 5C (Lieberman-Aiden *et al*, 2009). Once the technology has been developed to increase resolution, and with sufficient bioinformatic ability, Hi-C would be an ideal method to identify chromatin interactions in that were disrupted in CdLS cells compared to wildtype.

This data would be well complemented by data on cohesin and NIPBL genome interactions. It would be worthwhile to carry out ChIP or DamID experiments with cohesin and NIPBL, and to compare their genomic localisation with regions where chromatin interactions were disrupted in CdLS.

Previous studied have found that cohesin tends to localise near to the transcription start sites of genes misregulated in CdLS or upon NIPBL or SMC1 knockdown in that cell line (Liu *et al*, 2009; Kagey *et al*, 2010). A further

complement to interaction and localisation data would be studies of genes misregulated in CdLS in that tissue.

It would be important to carry out these studies in a number of cell lines and tissues, as it is likely that different interactions would occur and be affected differently in different tissues. If we were able to identify tissue-specific proteins that interact with cohesin, it would be interesting to compare all of these data

7.4 Conclusions

Without a consistent cellular phenotype between CdLS cell lines, it is difficult to make firm conclusions about the compaction of chromatin in CdLS. However, the fact that chromatin is decompacted in some CdLS cell lines and upon SMC1 and NIPBL knockdown, suggests that NIPBL and SMC1 do influence chromatin compaction, probably by altering chromosome topology. It is also likely that this occurs independently of CTCF, and differs between cell lines. It is likely this loss of cohesin-mediated intra-chromosome interactions that causes aberrant gene expression in CdLS.

8 References

8.1 Journals

Arumugam, P., Gruber, S., Tanaka, K., Haering, C.H., Mechtler, K. and Nasmyth, K., (2003) ATP hydrolysis is required for cohesin's association with chromosomes, *Current Biology*, **13**, 1941-1953

Bao, L., Zhou, M. and Cui, Y., (2008) CTCF BSDS: a CTCF-binding site database for characterisation of vertebrate genomic insulators, *Nucleic Acids Research*, **36**, D83-D87

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. And Zhao, K., (2007) High-resolution profiling of histone modifications in the Human genome, *Cell*, **129**, 823-837

Bermúdez-López, M., Ceschia, A., de Piccoli, G., Colomina, N., Pasero, P., Aragón, L. and Torres-Rosell, J., (2010) The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages, *Nucleic Acids Research*, Epub ahead of print

Bernard, P., Maure, J., Partridge, J.F., Genier, S., Javerzat, J. and Allshire, R.C., (2001) Requirement of heterochromatin for cohesion at centromeres, *Science*, **294**, 2539-2542

Bhiuyan, Z.A., Klein, M., Hammond, P., van Haeringen, A., Mannens, M.M., van Berckelaer-Onnes, I. And Hennekam, R.C. (2006) Genotype-phenotype correlations of 39 patients with Cornelia de Lange syndrome: the Dutch experience, *Journal of Medical Genetics*, **43**, 568-575

Biswas, M., Maquani, N., Rai, R., Kumaran, S.P., Iyer, K.R., Sendinc, E., Smith, J.S. and Laloraya, S., (2009) Limiting the extent of the *RDN1* heterochromatin domain by a silencing barrier and Sir2 protein levels in *Saccharomyces cerevisiae*, *Molecular Cell Biology*, **29**, 2889-2898

Bose, T. and Gerton, J.L., (2010) Cohesinopathies, gene expression and chromatin organisation, *Journal of Cell Biology*, **189**, 201-210

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N., Ellis, J.A. and Bickmore, W.A., (2001) The spatial organisation of human chromosomes within the nuclei of normal and emerin-mutant cells, *Human Molecular Genetics*, **10**, 211-219104-110.

Carruthers, L.M. and Hansen, J.C., (2000) The core histone N termini function independently of linker histones during chromosome condensation, *Journal of Biological Chemistry*, **275**, 37285-37290

Castronovo, P., Gervasini, C., Cereda, A., Masciadri, M., Milani, D., Russo, S., Selicorni, A. and Larizza, L., (2009) Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome, *Chromosome Research*, **17**, 763-71

Chambeyron, S. and Bickmore, W.A., (2004) Chromatin decondensation and nuclear reorganisation of the *HoxB* locus upon induction of transcription, *Genes and Development*, **18**, 1119-1130

Chan, R.C., Severson, A.F. and Meyer, B.J., (2004) Condensin restructures chromosomes in preparation for meiotic divisions, *Journal of Cell Biology*, **167**, 613-625

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y.W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S.Y., Chiang, C.M., Loukinov, D., Lobanenkov, V., Ohlsson, R. and Klenova, E., (2007) CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide, *Molecular Cell Biology*, **27**, 1631-1648

Cheung, V.G., Speilman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T., (2005) Mapping determinants of human gene expression by regional and genome-wide association, *Nature*, **437**, 1365-1369

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A. and Nasmyth, K., (2000) Cohesin's binding to chromosomes depends in a separate complex consisting of Scc2 and Scc4 proteins, *Molecular Cell*, **5**, 243-254

Cobbe, N., Savvidou, E. And Heck, M.M.S., (2006) Diverse mitotic and interphase functions of condensins in Drosophila, *Genetics*, **172**, 991-1008

Cohen-Fix, O., Peters, J.-M., Kirschner, M.W. and Koshland, D., (1996) Anaphase initiation in *Saccharomyces cerevisiae* is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p, *Genes and Development*, **10**, 3081-3093

Comings, D.E., (1967) Sex chromatin, nuclear size and the cell cycle, *Cytogenetic and Genome Research*, **6**, 120-144

Cremer, T. and Cremer, C., (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nature Reviews in Genetics*, **2**, 292-301

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P. and Bickmore, W.A., (1999) Differences in the localisation and morphology of chromosomes in the human nucleus, *Journal of Cell Biology*, **145**, 1119-1131

Csankovszki, G., (2009) Condensin function in dosage condensation, *Epigenetics*, **4**, 212-215

Cunningham, M.D., Kassis, J.A. and Pfeifer, K. (2010) Chromatin modifiers, cognitive disorders, and imprinted genes, *Developmental Cell*, **18**, 169-170

D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K. and Uhlmann, F., (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes, *Genes and Development*, **15**, 2215-2227

Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A. D., Wilson, M., Lillquist, K., Siu, V., Ramos, F. J., Musio, A., Jackson, L. S., Dorsett, D. and Krantz, I. D. (2007) Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation, *American Journal of Human Genetics*, **80**, 485-494

Degner, S.C., Wong, T.P., Jankevicius, G. and Feeney, A.J., (2009) Developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development, *Journal of Immunology*, **182**, 44-48

Dekker, J., Rippe, K., Dekker, M., Klekner, N., (2002) Capturing Chromosome Condensation, *Science*, **295**, 1306-1311

Del Campo, M., Jones, M.C., Veraksa, A.N., Curry, C.J., Jones, K.L., Mascarello, J.T., Ali-Kahn-Catts, Z., Drumheller, T. and McGinnis, W., (1999) Monodactylous limbs and abnormal genitalia are associated with hemizygosity for the human 2q13 regions that includes the HoxD locus, *American Journal of Human Genetics*, **65**,

Denes, V., Pilichowska, M., Marakaroskiy, A., Carpinito, G. and Geck, P., (2010) Loss of a cohesin-linked suppressor APRIN (Pds5b) disrupts stem cell programs in embryonal carcinoma: an emerging cohesin role in tumour suppression, *Oncogene*, **29**, 3446-3452

De Piccoli, G., Torres-Rosell, J. and Aragón, L., (2009) The unnamed complex: what do we know about Smc5-Smc6?, *Chromosome Research*, **17**, 251-263

Donze, D., Adams, C.R., Rine, J. and Kamakaka, R.T., (1999) The boundaries of the silenced HMR domain in *Saccharomyces cervevisiae*, *Genes and Development*, **13**, 698-708

Dorsett, D., Eissenberg, J.C., Misulovin, Z., Martens, A., Redding, B. and McKim, K., (2005) Effects of sister chromatid cohesion proteins on *cut* gene expression during wing development in *Drosophila*, *Development*, **132**, 4743-4753

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D. and Dekker, J., (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interaction between genomic elements, *Genome Research*, **16**, 1299-1309

Duan, X., Yang, Y., Chen, Y.-H., Arenz, J., Rangi, G.K., Zhao, X. and Ye, H., (2009) The architecture of the Smc5/6 complex of *S. cerevisiae* reveals a unique interaction between the Nse5-6 subcomplex and the hinge regions of Smc5 and Smc6, *Journal of Biological Chemistry*, **284**, 8507-8515

Dunn, K.L., Zhao, H. and Davie, J.R., (2003) The insulator binding protein CTCF associates with the nuclear matrix, *Experimental Cell Research*, **288**, 218-223

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A. and Bickmore, W.A., (2010) Ring1B compacts

chromatin structure and represses gene expression independent of histone ubiquitination, *Molecular Cell*, **38**, 452-464

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A. and Klenova, E., (2010) Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides and insight into the mechanism of its regulation by pol(ADP-ribosyl)ation, *Molecular Cell Biology*, **30**, 1199-1216

Fazzio, T.G. and Panning, B., (2010) Condensin complexes regulate mitotic progression and interphase chromatin structure in embryonic stem cells, *Journal of Cell Biology*, **188**, 491-503

Furuya, K., Takahashi, K. and Yanagida, M., (1998) Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G_1 phase, *Genes and Development*, **12**, 3408-3418

Gard, S., Light, W., Xiong, B., Bose, T., McNairn, A/B., Harris, B., Fleharty, B., Seidel, C., Brickner, J.H. and Gerton, J.L., (2009) Cohesinopathy mutations disrupt the subnuclear organisation of chromatin, *Journal of Cell Biology*, **187**, 455-462

Gause, M., Misulovin, Z., Bileu, A. and Dorsett, D., (2010) Dosage sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5 and Wapl, *Molecular and Cellular Biology*, Epub ahead of print

Gérard, M., Duboule, D. and Zákány, J., (1993) Cooperation of regulatory elements involved in the activation of the Hoxd-11 gene, *Comptes Rendus de l'Académie des Sciences*, **316**, 985-994

Gerlich, D., Koch, B., Dupeux, F., Peters, J.M. and Ellenberg, J., (2006) Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication, *Current Biology*, **16**, 1571-1578

German, J., (1979) Roberts' syndrome. I. Cytological evidence for a disturbance in chromatid pairing, *Clinical Genetics*, **16**, 441-447

Ghiselli, G., (2006) SMC3 knockdown triggers genomic instability and p53-dependant apoptosis in human and zebrafish cells, *Molecular Cancer*, **5**, 52

Ghosh, S.K., Huang, C.-C., Hajra, S. and Jayaram, M., (2010) Yeast cohesin complex embrace 2 micron plasmid sisters in a trilinked catenane complex, *Nucleic Acids Research*, **38**, 570-584

Gierman, H.J., Indemans, M.H., Koster, J., Goetze, S., Seppen, J., Geerts, D., van Driel, R. and Versteeg, R., (2007) Domain-wide regulation of gene expression in the human genome, *Genome Research*, **17**, 1286-1295

Gilbert, N. and Allan, J., (2001) Distinctive higher-order chromatin structure at mammalian centromeres, *Proceedings of the National Academy of Sciences*, **98**, 11949-11954

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A., (2004) Chromatin architecture of the human genome: gene rich domains are enriched in open chromatin fibres, *Cell*, **118**, 555-566 binding but not chromatin compaction, *Journal of Cell Biology*, **177**, 401-411

Gillespie, P.J. and Hirano, T., (2004) Scc2 couples replication licensing to sister chromatid cohesion in *Xenopus* egg extracts, *Current Biology*, **14**, 1598-1603

Gillis, L.A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A.D., Li, H.H., Devoto, M., Jackson, L.G. and Krantz, I.D., (2004) NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations, *American Journal of Human Genetics*, **74**, 610-623

Goetze, S., Mateos-Langerak, J., Gierman, H.J., de Leeuw, W., Giromus, O., Indemans, M.H., Koster, J., Ondrej, V., Versteeq, R. and van Driel, R., (2007) The three-dimensional structure of human interphase chromosomes is related to the transcriptome map, *Molecular Cell Biology*, **27**, 4475-4487

Gomes, N.P. and Espinosa, J.M., (2010) Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding, *Genes and Development*, **24**, 1022-1034

Gordillo, M., Vega, H., Trainer, A.H., Hou, F., Sakai, N., Luque, R., Kayserili, H., Basaran, S., Skovby, F., Hennekam, R.C., Uzielli, M.L., Schnur, R.E., Manouvrier, S., Chang, S., Blair, E., Hurst, J.A., Forzano, F., Meins, M., Simola, K.O., Raas-Rothschild, A., Schultz, R.A., McDaniel, L.D., Ozono, K., Inui, K., Zou, H. and Jabs, E.W., (2008) The molecular mechanism underlying Roberts syndrome involves loss of ESCO2 acetyltransferase activity, *Human Molecular Genetics*, **17**, 2172-2180

Gosling, K.M., Mkaroff, L.E., Theodoratos, A., Kim, Y.-H., Whittle, B., Rui, L., Wu, H., Hong, N.A., Kennedy, G.C., Fritz, J.-A., Yates, A.L., Goodnow, C.C. and Fahrer, A.M., (2007) A mutation in a chromosome condensin II subunit, kleisin β, specifically disrupts T cell development, *Proceedings of the National Academy of Sciences*, **104**, 12445-12450

Grigoryev, S.A., Solovieva, V.O., Spirin, K.S. and Krasheninnikov, I.A., (1992) A novel nonhistone protein (MENT) promotoes nuclear collapse at the terminal stage of avian erythropoiesis, *Experimental Cell Research*, **198**, 268-275

Gullerova, M. and Proudfoot, N.J., (2008) Cohesin complex promotes transcription termination between convergent genes in *S. pombe*. *Cell*, **132**, 983-995

Gurdon, J.B., (1976) Injected nuclei in frog oocytes: fate, enlargement, and chromatin dispersal, *Journal of Embryology and Experimental Morphology*, **36**, 523-540

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G. and Merkenschlager, M. (2009) Cohesin form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus, *Nature*, **460**, 410-413

Haering, C.H., Farcas, A-M, Arumugam, P., Metson, J. and Nasmyth, K., (2008) The cohesin ring concatenates sister DNA molecules, *Nature*, **454**, 297-301

Hagstrom, K.A., Holmes, V.F., Cozzarelli, N.R. and Meyer, B.J., (2002) *C. elegans* condensin promotes mitotic chromosome architecture, centromere organisation, and sister chromatid segregation during mitosis and meiosis, *Genes and Development*, **16**, 729-742

Hakimi, M.-A., Bochar, D.A., Schmiesing, J.A., Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K. and Shiekhattar, R., (2002) A chromatin remodelling complex that loads cohesin onto human chromosomes, *Nature*, **418**, 994-998

Hallson, G., Syrzycka, M., Beck, S.A., Kennison, J.A., Dorsett, D., Page, S.L., Hunter, S.M., Keall, R., Warren, W.D., Brock, H.W., Sinclair, D.A. and Honda, B.M. (2008) The *Drosophila* cohesin subunit Rad21 is a trithorax group (trxG) protein, *Proceedings of the National Academy of Sciences USA*, **105**, 12405-12410

Hartl, T.A., Sweeney, S.J., Knepler, P.J. and Bosco, G., (2008) Condensin II resolves chromosomal associations to enable anaphase I segregation in *Drosophila* male meiosis, *PLoS Genetics*, **4**, e1000228

Heard, E. and Bickmore, W.A., (2007) The ins and outs of gene regulation and chromosome territory organisation, *Current Opinions in Cell Biology*, **19**, 311-316

Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., can de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F., Hendricks, R.W. and Galjart, N. (2008) CTCF regulates cell cycle progression of alphabeta T cells in the thymus, *EMBO Journal*, **27**, 2839-2850

Hirano, T., Kobayashi, R. And Hirano, M., (1997) Condensins, chromosome condensation complexes containing XCAP-C, XCAP-E and a Xenopus homologue of the Drosophila Barren protein, *Cell*, 511-521

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schübeler, D. and Gilbert, D.M., (2008) Global reorganisation of replication domains during embryonic stem cell differentiation, *PLoS Biology*, **6**, e245

Horsfield, J.A., Anagnostou, S.H., Hu, J.K-H, Cho, K.H.Y., Geisler, R., Lieschke. G., Crosier, K.E. and Crosier, P.S., (2007) Cohesin-dependant regulation of Runx genes, *Development*, **134**, 2639-2649

Hou, C., Dale, R. and Dean, A., (2010) Cell type specificity of chromatin organisation mediated by CTCF and cohesin, *Proceedings of the National Academy of Sciences*, **23**, 3651-3656

Huang, J., Hsu, J.M. and Laurent, B.C., (2004) The RSC nucleosome-remodelling complex is required for cohesin's association with chromosome arms, *Molecular Cell*, **13**, 739-750

Hudson, D.F., Vagnarelli, P., Gassman, R. and Earnshaw, W.C., (2003) Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes, *Developmental Cell*, **5**, 323-336

Ishihara, K., Oshimura, M. and Nakao, M., (2006) CTCF-dependant chromatin insulator is linked to epigenetic remodelling, *Molecular Cell*. **23**, 733-742

Ivanov, D. and Nasmyth, K., (2007) A physical assay for sister chromatid cohesion *in vitro*, *Molecular Cell*, **27**, 300-310

Jabs, E.W., Tuck-Muller, C.M., Cusano, R. and Rattner, J.B., (1991) Studies of mitotic and centromeric abnormalities in Roberts syndrome: implications for a defect in the mitotic mechanism, *Chromosoma*, **100**, 251-261Jack, J., Dorsett, D., Delotto, Y. and Liu, S., (1991) Expression of the *cut* locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer, *Development*, **113**, 735-747

Jahnke, P., Xu, W., Wülling, M., Albrecht, M., Gabriel, H., Gillessen-Kaesbach, G. and Kaiser, F.J., (2008) The cohesin loading factor NIPBL recruits histone deacetylases to mediate local chromatin modifications, *Nucleic Acids Research*, **36**, 6450-6458

Johnson, T.S., Swartzendruber, D.E. and Martin, J.C., (1981) Nuclear size of G1/S transition cells measured by flow cytometry, *Experimental Cell Research*, **134**, 201-205

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J. and Young, R.A., (2010) Mediator and cohesin connect gene expression and chromatin architecture, *Nature*, Epub ahead of print

Kang, H. and Lieberman, P.M., (2009) Cell cycle control of KSHV latency transcription by CTCF-cohesin interactions, *Journal of Virology*, **83**, 6199-6210

Kang, K., Lee, S.B., Yoo, J.H and Nho, C.W., (2010) Flow cytometric fluorescence pulse width analysis of etopiside-induced nuclear enlargement in HCT116 cells, *Biotechnology Letters*, **32**, 1045-1052

Kaur, M., DeScipio, C., McCallum, J., Yaeger, D., Devoto, M., Jackson, L.G., Spinner, N.B. and Krantz, I.D., (2005) Precocious sister chromatid separation (PSCS) in Cornelia de Lange syndrome, *American Journal of Medical Genetics*, **138**, 27-31

Kawauchi, S., Calof, A.L., Santos, R., Lopez-Burks, M.E., Young, C.M., Hoang, M.P., Chua, A., Lao, T., Lechner, M.S., Daniel, J.A., Nussenzweig, A., Kitzes, L., Yokomori, K., Hallgrimsson, B. and Lander, A.D., (2009) Multiple organ system defects and transcriptional dysregulation in the *Nipbl*$^{+/-}$ mouse, a model of Cornelia de Lange Syndrome, *PLoS Genetics*, **5**, e1000650

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D., (2002) The human genome browser at UCSC, *Genome* Research, **12**, 996-1006

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R.W. and Bérubé, N.G., (2010) ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain, *Developmental Cell*, **18**, 191-201

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V. and Ren, B., (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the Human genome, *Cell*, **128**, 1231-1245

Kimura, K., Rybenkov, V.V., Crisona, N.J., Hirano, T. and Cozzarelli, N.R., (1999) 13S condensin actively reconfigures DNA by introducing global positive writhe : implications for chromosome condensation, *Cell*, **98**, 239-248

Kline, A.D., Grados, M., Sponseller, P., Levy, H.P., Blagowidow, N., Schoedel, C., Rampolla, J., Clemens, D.K., Krantz, I., Kimball, A., Pichard, C., Tuchman, D., (2007) Natural history of aging in Cornelia de Lange syndrome, *American Journal of Medical Genetics Part C (Seminars in Medical Genetics)*, **145**, 248-260

Kogut, I., Wang, J., Guacci, V., Mistry, R.K. and Megee, P.C., (2009) The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes, *Genes and Development*, **23**, 2345-2357

Kools, P., Van Imschoot, G. and van Roy, F., (2000) Characterisation of three novel human cadherin genes (CDH7, CDH19, and CDH20) clustered in chromosome 18q22-q23 and with high homology to chicken cadherin-7, *Genomics*, **68**, 283-295

Krantz, I.D., McCallum, J., DeSpicio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A., Nowaczyk, M.J.M., Toriello, H., Bamshad, M.J., Carey, J.C., Rappaport, E., Kawauchi, S., Lander, A.D., Calof, A.L., Li, H.H., Devoto, M. and Jackson, L.G., (2004) Cornelia de Lange syndrome is caused by mutations in *NIPBL*, the human homolog of *Drosophila melanogaster Nipped-B*, *Nature Genetics*, **36**, 631-635

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K. and Peters, J.M., (2006) Wapl controls the dynamic association of cohesin with chromatin, *Cell*, **127**, 955-967

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W. and Ohlsson, R., (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2, *Proceedings of the National Academy of Sciences USA*, **103**, 10684-10689

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G. and Rauscher, F.J. (2005) The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain, *Biomedical and Biophysical Research Communications*, **331**, 929-937

Lee, K., Kang, M.J., Kwon, S.J., Kwon, Y.K., Kim, K.W., Lim, J.H. and Kwon, H., (2007) Expansion of chromosome territories with chromatin decompaction in BAF53-depleted interphase cells, *Molecular Biology of the Cell*, **18**, 4013-4023

Legronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K. and Uhlmann, F., (2004) Cohesin relocalisation from sites of chromosomal loading to places of convergent transcription, *Nature*, **430**, 573-578

Lehnhardt, M., Klein-Hitpass, L., Kuhnen, C., Homann, H.H., Daigeler, A., Steinau, H.U., Roehrs, S., Schnoor, L., Steinstraesser, L. and Mueller, O., (2005) Response rate of fibrosarcoma cells to cytotoxic drugs on the expression level correlates to the theraputic response rate of fibrosarcomas and is mediated by regulation of apoptotic pathways, *BioMed Central Cancer*, **5**, 74

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S. and Dekker, J., (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science*, **326**, 289-293

Lim, J.-H. And Oh, B.-H., (2009) Structural and functional similarities between two bacterial chromosome compacting machineries, *Biochemical and Biophysical Research Communications*, **386**, 415-419

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N.B., Vega, H., Jackson, L.G., Shirahige, K. and Krantz, I.D., (2009) Transcriptional dysregulation of *NIPBL* and cohesin mutant human cells, *PLoS Biology*, **7**, e1000119

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Li, J.R., Clark, D., Kair, M., Tatsuro, K., Kline, A.D., Chang, C., Vega, H., Jackson, L.D., Spinner, N.B., Shirahige, K and Krantz, I.D., (2010) Genome-wide DNA methylation analysis in cohesin mutant cell lines, *Nucleic Acids Research*, Epub ahead of print

Losada, A., Hirano, M. and Hirano, T., (1998) Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion, *Genes and Development*, **12**, 1986-1997

Losada, A. and Hirano, T., (2005) Dynamic molecular linkers of the genome: the first decade of SMC proteins, *Genes and Development*, **19**, 1269-1287

Lupo, R., Breiling, A., Bianchi, M.E. and Orlando, V., (2001) *Drosophila* chromosome condensation proteins topoisomerase II and barren colocalise with polycomb and maintain *Fab-7* PRE silencing, *Molecular Cell*, **7**, 127-136

MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J. and MacAlpine, D.M., (2010) *Drosophila* ORC localises to open chromatin and marks sites of cohesin complex loading, *Genome Research*, **20,** 201-211

Mahy, N.L., Perry, P.E., Gilchrist, S., Baldock, R.A. and Bickmore, W.A., (2002) Spatial organisation of active and inactive genes and noncoding DNA within chromosome territories, *Journal of Cell Biology*, **157**, 579-589

Mahy, N.L., Perry, P.E. and Bickmore, W.A., (2002) Gene density and transcription influence the localisation of chromatin outside of chromosome territories detectable by FISH, *Journal of Cell Biology*, **159**, 753-763

Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E.M., Verschure, P.J., Indemans, M.H., Gierman, H.J., Heermann, D.W., van Driel, R. and Goetze, S., (2009) Spatially confined folding of chromatin in the interphase nucleus, *Proceedings of the National Academy of Sciences*, **106**, 3812-3817

McNairn, A.J. and Gerton, J.L., (2009) Intersection of ChIP and FLIP, genomic methods to study the dynamics of cohesin proteins, *Chromosome Research*, **17**, 155-163

Michaelis, C., Ciosk, R., Nasmyth, K., (1997) Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids, *Cell*, **91**, 35-45

Mishuro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y. and Nakao, M., (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster, *The EMBO Journal*, **28**, 1234-1245

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D. and Dorsett, D., (2008) Association of cohesin and Nipped-B with transcriptionally active regions of the *Drosophila melanogaster* genome, *Chromosoma*, **117**, 89-102

Mitchell, J.A. and Fraser, P., (2008) Transcription factories are nuclear subcompartments that remain in the absense of transcription, *Genes and Development*, **22**, 20-25

Mönnich, M., Banks, S., Eccles, M., Dickinson, E. and Horsfield, J., (2009) Expression of cohesin and condensin genes during zebrafish development supports and non-proliferative role for cohesin, *Gene Expression Patterns*, **9**, 586-594

Morey, C., Da Silva, N.R., Perry, P. and Bickmore, W.A., (2007) Nuclear reorganisation and chromatin decondensation are conserved but distinct, mechanisms linked to Hox gene activation, *Development*, **134**, 909-919

Morey, C., Kress, C. and Bickmore, W.A., (2009) Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression, *Genome Research*, **19**, 1184-1194

Müller, I., Boyle, S., Singer, R.H., Bickmore, W.A. and Chubb, J.R., (2010) Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells, *PLoS One*, **13**, e11560

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P. and Larizza, L. (2006) X-linked Cornelia de Lange syndrome owing to *SMC1L1* mutations**,** *Nature Genetics,* **38,** 528-530

Nasmyth, K., (1999) Separating sister chromatids, *Trends in Biochemical Sciences*, **24**, 98-104

Nasmyth, K. and Schleiffer, A., (2003) From a single double helix to paired double helices and back, *Philosophical Transactions of the Royal Society of London*, **359**, 99-108

Nasmyth, K., (2005) How might cohesin hold sister chromatids together?, *Philosophical Transactions of the Royal Society of London*, **360**, 483-496

Nasmyth, K. and Haering, C., (2009) Cohesin: Its roles and mechanisms, *Annual Review of Genetics*, **43**, 525-558

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.-M. and Murrell, A., (2009) Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus, *PloS Genetics*, **5,** e1000739

Neuwald, A.F. and Hirano, T., (2000) HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions, *Genome Research*, **10**, 1445-1452

Ng, S.B., Nickerson, D.A., Bamshad, M.J. and Shendure, J., (2010) Massively parallel sequencing and rare disease, *Human Molecular Genetics*, **19**, R119-R124

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G. and Merkenschlager, M. (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms, *Cell*, **132**, 422-433

Parkhurst, S.M. and Corces, V.G., (1986) Interactions among the gypsy transposable element and the yellow and the suppressor of hairy-wing loci in *Drosophila melanogastor*, *Molecular and Cellular Biology*, **6**, 47-53

Pauli, A., Althoff, F., Oliveira, R.A., Heidmann, S., Schuldiner, O., Lehner, C.F., Dickson, B.J. and Nasmyth, K., (2008) Cell-type-specific TEV protease cleavage reveals cohesin functions in *Drosophila* neurons, *Developmental Cell*, **14**, 239-251

Pedrosa, E., Stefanescu, R., Margolis, B., Petruolo, O., Lo, Y., Nolan, K., Noval, T., Stopkova, P. and Lachman, H.M., (2008) Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia, *Schizophrenia Research*, **102**, 210-219

Peters, J.-M., Tedeschi, A. and Schmitz, J., (2008) The cohesin complex and its roles in chromosome biology, *Genes and Development*, **22**, 3089-3114

Rasheed, S., Nelson-Rees, W.A., Toth, E.M., Arnstein, P. and Gardner, M.B., (1974) Characterisation of a newly derived human sarcoma cell line (HT1080), *Cancer*, **33**, 1027-1033

Reeves, J., Wilson, D., Sharkey, F. and FitzPatrick, D.R., (in preparation) Deletion of NIPBL detected on array-CGH associated with atypical Cornelia de Lange Syndrome

Revenkova, E., Focarelli, M.L., Susani, L., Paulis, M., Bassi, M.T., Mannini, L., Frattini, A., Delia, D., Krantz, I., Vezzoni, P., Jessberger, R. and Musio, A., (2009)

Cornelia de Lange Syndrome mutations on SMC1A or SMC3 affect binding to DNA, *Human Molecular Genetics*, **18**, 418-427

Reyes, E.D., Patidar, P.L., Uranga, L.A., Bortoletto, A.S. and Lusetti, A.L., (2010) RecN is a cohesin-like protein that stimulates intermolecular DNA interaction *in vitro*, *Journal of Biological Chemistry*, **285**, 16521-16529

Ribich, S., Tasic, B. and Maniatis, T., (2006) Identification of long-range regulatory elements in the protocadherin-α gene cluster, *Proceedings of the National Academy of Sciences*, **103**, 19719-19724

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G. and Rhodes, D., (2008) 30nm chromarin fibre decompaction requires both H4-K16 acetylation and linker histone eviction, *Journal of Molecular Biology*, **381**, 816-825

Rollins, R.A., Morcillo, P. and Dorsett, D., (1999) Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the *cut* and *Ultrabithorax* genes, *Genetics*, **152**, 577-593

Rollins, R.A., Korom, M., Aulner, N., Martens, A. and Dorsett, D., (2004) *Drosophila* Nipped-B protein supports sister chromatid cohesion and opposes the Stromalin/Scc3 cohesion factor to facilitate long-range activation of the *cut* gene, *Molecular and Cellular Biology*, **24**, 3100-3111

Rowland, B.D., Roig, M.B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouët, F., Underwood, P., Metson, J., Imre, R., Mechtler, K., Katis, V.L. and Nasmyth, K., (2009) Building sister chromatid cohesion: Smc3 acetylation counteracts an antiestablishment activity, *Molecular Cell*, **33**, 763-774

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filipova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A. and Krumm, A., (2008) CTCF physically links cohesin to chromatin, *Proceedings of the National Academy of Science*, **105**, 8309-8314

Sachs, R.K., van der Engh, G., Trask, B., Yokota, H. and Hearst, J.E. (1995) A random-walk/giant-loop model for interphase chromosomes, *Proceedings of the National Academy of Sciences*, **92**, 2710-2714

Schaaf, C.A., Misulovin, Z., Sahota, G., Siddiqui, A.M., Schwartz, Y.B., Kahn, T.G., Pirrotta, V., Gause, M. and Dorsett, D., (2009) Regulation of the *Drosophila*

Enhancer of split and *invected-engrailed* gene complexes by sister chromatid cohesion proteins, *PLoS ONE*, **4**, e6202

Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K. and Eisenhaber, F., (2003) Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners, *Molecular Cell*, **11**, 571-575

Schmidt, D., Schwalie, P.C., Ross-Innes, C., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek and Odom, D.T., (2010) A CTCF-independent role for cohesin in tissue specific transcription, *Genome Research*, **20**, 578-588

Schmiedeberg, L., Skene, P., Deaton, A. and Bird, A., (2009) A temporal threshold for formaldehyde crosslinking and fixation, *PLoS ONE*, **4**, e4636

Schuldiner, O., Berdnik, D., Levy, J.M., Wu, J.S., Luginbuhl, D., Gontang, A.C. and Luo, L., (2008) *piggy*Bac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning, *Developmental Cell*, **14**, 227-238

Seitan, V.C., Banks, P., Laval, S., Majid, N.A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., Rollins, R.A., Erdjument-Bromage, H., Tempst, P., Benard, C.Y., Hekimi, S., Newbury, S.F. and Strachan, T., (2006) Metazoan Scc4 homologues link sister chromatid cohesin to cell and axon migration guidance, *PLoS Biology*, **4**, 1411-1425

Selicorni, A., Russo, S., Gervasini, C., Castronovo, P., Milani, D., Cavalleri, F., Bentivegna, A., Masciadri, M., Domi, A., Divizia, M.T., Sforzini, C., Tarantino, E., Memo, L., Scarano, G and Larizza, L., (2007) Clinical score of 62 patients with Cornelia de Lange syndrome and correlations with the presense and type of NIPBL mutation, *Clinical Genetics*, **72** Selicorni, 98-108

Serrano, A., Rodríguez-Corsino, M. and Losada, A., (2009) Heterochromatin Protein 1 (HP1) proteins do not drive pericentromeric cohesin enrichment in human cells, *PLoS ONE*, **4**, e5118

Sharpless, T.K. and Melamed, M.R., (1976) Estimation of cell size from pulse shape in flow cytofluorometry, *Journal of Histochemistry and Cytochemistry*, **24**, 257-264

Shintomi, K. and Hirano, T., (2009) Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1, *Genes and Development*, **23**, 2224-2236

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R. and Peterson, C.L., (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions, *Science*, **311**, 844-847

Sipos, L., and Gyurkovics, H., (2005) Long distance interactions between enhancers and promoters, *The FEBS Journal*, **272**, 3253-3259

Skibbens, R.V., Marzillier, J. and Eastman, L., (2010) Cohesins coordinate gene transcriptions of related function in *Saccharomyces cerevisiae*, **9**, 1601-1606

Splinter, E., Heath, H., Kooren, J., Palstra, R-J., Klous, P., Grosweld, F., Galjart, N. and de Laat, W., (2006) CTCF mediates long range chromosome looping and local histone modification in the β-globin locus, *Genes and Development*, **20**, 2349-2354

Stedman, W., Kang, H., Lin, S., Kissil, J., Bartolomei, M.S. and Lieberman, P.M., (2008) Cohesins localise with CTCF at the HSV latency control region and at cellular c-myc and H19/Igf2 insulators, *The EMBO Journal*, **27**, 654-666

Strunnikov, A.V., Larionov, V.L and Koshland, D., (1993) *SMC1*: An essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family, *Journal of Cell Biology*, **6**, 1635-1648

Strunnikov, A.V., (2006) SMC complexes in bacterial chromosome condensation and segregation, *Plasmid*, **55**, 135-144

Sutherland, H. and Bickmore, W.A., (2009) Transcription factories: gene expression in unions?, *Nature Reviews Genetics*, **10**, 457-466

Taatjes, D.J., (2010) The human Mediator complex: a versatile, genome-wide regulator of transcription, *Trends in Biochemical Sciences*, **35**, 315-322

Tagami, T., Hirose, K., Barichello, J.M., Ishida, T. and Kiwada, H., (2008) Global gene expression profiling in cultured cells is strongly influenced by treatments with siRNA-cationic lipsome complexes, *Pharmaceutical Research*, **25**, 2497-2504

Takahashi, T.S., Yiu, P., Chou, M.F., Gygi, S. and Walter, J.C., (2004) Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex, *Nature Cell Biology*, **6**, 991-996

Tanaka, K., Hao, Z., Kai, M. and Okayama, H., (2001) Establishment and maintenance of sister chromatid cohesion in fission yeast by a unique mechanism, *The EMBO Journal*, **20**, 5779-5790

Terret, M.E., Sherwood, R., Rahman, S., Qin, J. and Jallepalli, P.V., (2009) Cohesin acetylation speeds the replication fork, *Nature*, **462**, 231-234

Tiwari, V. K., Cope, L., McGarvey, K.M., Ohm, J.E. and Baylin, S.B. (2008) A novel 6C assay uncovers polycomb-mediated higher order chromatin conformations, *Genome Research*, **18**, 1171-1179

Tomkins, D.J. and Sisken, J.E., (1984) Abnormalities in the cell-division cycle in Roberts syndrome fibroblasts: a cellular basis for the phenotypic characteristics?, *American Journal of Human Genetics*, **36**, 1332-1340

Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J. and Strachan, T., (2004) *NIPBL*, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome, *Nature Genetics*, **36**, 636-641

Tóth, K.F., Knoch, T.A., Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C.P., Müller, G. and Rippe, K., (2004) Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin, *Journal of Cell Science*, **117**, 4277-4287

Tse, C., Sera, T., Wolffe, A.P. and Hansen, J.C., (1998) Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III, *Molecular Cell Biology*, **18**, 4629-4638

Uhlmann, F. and Nasmyth, K., (1998) Cohesin between sister chromatids must be established during DNA replication, *Current Biology*, **8**, 1095-1101

Ulhmann, F., Wernic, D., Poupart, M.-A., Koonin, E.V. and Nasmyth, K., (2000), Cleavage of cohesin by the CD clan protease Separin triggers anaphase in yeast, *Cell*, **103**, 375-386

Van den Berg, D.J., Francke, U., (1993) Sensitivity of Roberts syndrome cells to gamma radiation, mitomycin C, and protein synthesis inhibitors, *Somatic Cell Molecular Genetics*, **9**, 377-392

Van den Engh, G., Sachs, R. and Trask, B., (1992) Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model, *Science*, **257**, 1410-1412

Van der Lelij, P., Godthelp, B.C., van Zon, W., van Gosliga, D., Oostra, A.B., Steltenpool, J., de Groot, J., Scheper, R.J., Wolthuis, R.M., Waisfisz, Q., Darroudi, F., Joenje, H. and de Winter, J.P., (2009) The cellular phenotype of Roberts Syndrome fibroblasts as revealed by ectopic expression of ESCO2, *PloS ONE*, **4**, e6936

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., Jabs, E.W., Inui, K. and Joenje, H., (2005) Roberts syndrome is caused by mutations in *ESCO2*, a human homolog of yeast *ECO1* that is essential for the establishment of sister chromatid cohesion, *Nature Genetics*, **37**, 468-470

Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K and van Steensel, B., (2006) Human heterochromatin proteins form large domains containing KRAB-ZNF genes, *Genome Research*, **16**, 1493-1504

Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., Trowsdale, J. and Sheer, D., (2000) Large-scale chromatin organisation of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei, *Journal of Cell Science*, **113**, 1565-1576

Waizenegger, I.C., Hauf, S., Meinke, A. and Peters, J.M., (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase, *Cell*, **103**, 399-410

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K. and Peters, J.-M., (2006) Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression, *Current Biology*, **16**, 863-874

Watrin, E. and Peters, J.-M., (2009) The cohesin complex is required for the DNA damage-induced G_2/M checkpoint in mammalian cells, *The EMBO Journal*, **28**, 2625-2635

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J., (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor, *Nature*, **451**, 796-803

Williams, R.R., Broad, S., Sheer, D. and Ragoussis, J., (2002) Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoplast interphase nuclei, *Experimental Cell Research*, **272**, 163-175

Wolffe, A.P. and LeBlanc, B.P., (2000) Creating molecular clues to uncover gene function, *Nature Biotechnology*, **18**, 379-380

Wong, R.W. and Blobel, G., (2008) Cohesin subunit SMC1 associates with mitotic microtubules at the spindle pole, *Proceedings of the National Academy of Sciences*, **105**, 15441-15445

Wood, A.J., Severson, A.F. and Meyer, B.J., (2010) Cohesin and condensin complexity: the expanding repertoire of functions, *Nature Reviews Genetics*, **11**, 391-404

Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I. and Carter, N.P., (2004) Replication timing of the human genome, *Human Molecular Genetics*, **13**, 191-202

Yagi, T, (2008) Clustered protocadherin family, *Development, Growth and Differentiation*, **50**, 131-140

Yan, J., Zhang, F., Brundage, E., Scheurele, A., Lanpher, B., Erickson, R.P., Powis, Z., Robinson, H.B., Trapane, P.L., Stachiw-Hietpas, D., Keppler-Noreuil, K.M., Lalani, S.R., Sahoo, T., Chinault, A.C., Patel, A., Cheung, S.W. and Lupski, J.R, (2009) Genome duplication resulting in increased copy number of genes encoding

the sister chromatid cohesion complex conveys clinical consequences distinct from Cornelia de Lange, *Journal of Medical Genetics*, **46**, 626-634

Yokota, H., van der Engh, G., Hearst, J.E., Sachs, R.K. and Trask, B.J., (1995) Evidence for the organisation of chromatin in megabase pair-sized loops arranged along a random walk path in the human G_0/G_1 interphase nucleus, *Journal of Cell Biology*, **130**, 1239-1249

Yokota, H., Singer, M.J., van der Engh, G.J. and Trask, B.J. (1997) Regional difference in the compaction of chromatin in human G_0/G_1 interphase nuclei, *Chromosome Research*, **5**, 157-166

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., Oshimura, M., Feinberg, AP., Lobanenkov, V., Klenova, E. and Ohlsson, R., (2004) Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation, *Nature Genetics*, **36**, 1036-1037

Yusufzai, T.M., Tagami, H., Nakatani, Y. and Felsenfeld, G., (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species, *Molecular Cell*, **13**, 291-298

Zhang, B., Chang, J., Fu, M., Huang, J., Kashyap, R., Salavaggione, E., Jain, S., Shashikant, K., Deardorff, M.A., Giovannucci Uzielli, M.L., Dorsett, D., Beebe, D.C., Jay, P.Y., Heuckeroth, R.O., Krantz, I. and Millbrandt, J. (2009) Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies, *PloS ONE*, **4**, e5232

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R.O., Erlich, J.M., Jay, P.Y. and Milbrandt, J., (2007) Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome, *Development and Disease*, **134**, 3191-3201

Zhang, J., Shi, X., Li, Y., Kim, B.J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., Zhang, P., Kim, S.T., Pan, X. and Qin, J., (2008) Acetylation of Smv3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast, *Molecular Cell*, **31**, 143-151

Zhang, N., Kuznetsov, S.G., Sharan, S.K., Li, K., Rao, P.H. and Pati, D., (2008) A handcuff model for the cohesin complex, *Journal of Cell Biology*, **183**, 1019-1031

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh Sandhu, K., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. and Ohlsson, R., (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions, *Nature Genetics*, **38**, 1341-1347

Zou, H., McGarry, T.J., Bernal, T. and Kirshner, M.W., (1999) Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis, *Science*, **285**, 344-345

8.2 Websites

BacPac Resources Centre. Children's Hospital Oakland Research Institute, CA. : {27/9/07-14/5/09} : . World Wide Web URL: http://bacpac.chori.org/

Biomart. Ontario Institute for Cancer Research and European Bioinformatics Institute. : {6/9/10} : . World Wide Web URL: http://www.biomart.org/

Flowjo Flow Cytometry Analysis Software. Treestar, Inc. : {26/8/10} : . World Wide Web URL: http://www.flowjo.com/

Graphpad Software Inc. : {5/3/09, 4/11/09, 16/11/09} : . World Wide Web URL: http://www.graphpad.com/quickcalcs/pvalue1.cfm

Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {122470, 300590, 610759, 268300}: {13/2/08}: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

Welcome Trust Sanger Centre. Cambridge. : {12/4/07}: . World Wide Web URL: http://www.sanger.ac.uk/

UCSC Genome Browser. University of California, Santa Cruz, CA. Human genome GRCh37/h19 assembly : . World Wide Web URL: http://genome.ucsc.edu/

Acknowledgements

There are a lot of people who I have to thank both scientifically and personally for help during my PhD, so I'll start with the more serious thank-yous. I'd like to thank Professor Tom Strachan and Dr Matt Deardorff who sent me cell lines, and then sent me more when I managed to kill them. Many thanks to the people who gave me bioinformatics help, Graeme Grimes for helping me with UCSC tracks, Gogo for doing my bioinformatic comparisons which make up Chapter 6 and Emily Chambers for looking at my regions of interest. Thank you to Niall Anderson who helped me to choose the right statistics and wrote me a Macro to make Minitab do them for me, also thanks to Julia Dorin for suggesting speaking to a statistician and Patricia Yeyati for recommending a good one. Thanks to Craig Nicol for help with printing. Thanks to Technical Services at the HGU for all the things they do, making solutions, cleaning up and sorting out cells. Thanks to computing for fixing and installing things. And thanks to Paul Perry (with later help from Matt Pearson) for helping me with his expensive toys.

I have to thank Wendy for giving me the chance to do a PhD in her lab, for scientific guidance and most importantly for taking me to Italy. She is a brilliant scientist and an inspiration. Also, thanks to David Fitzpatrick for always being there with more ideas, and for giving me the chance to meet CdLS patients and put my work into context.

Any work in the Bickmore lab would be impossible without the two amazing Research Assistants, Shelagh and Liz, who are both just so knowledgable and so helpful that I know I would have gotten nothing done without them. Shelagh's FISH witch-craft is something to behold. I've been helped by all members of the lab, past and present, by asking for advice or stealing their reagents: Duncan, Catherine, Heidi, Ragnhild, Nishal, Iain, Gillian, Sehrish, Pradeepa, Pierre, Emmanuelle, Leisha and Rob. Also, thanks to Nick Gilbert for being on hand for the odd piece of advice.

Everybody on E3 has at some point offered me some help, especially Mary Taggart who holds the Section together.

And now to personal thanks. My family have always been there when things are getting me down, trying to listen to me grumble about a world they knew nothing about. Tom has taken good care of me over the past two years and I am so grateful to him. I have to thank my girls for chatting nonsense at lunchtime, Jenny, Corina, Natalie, Bethan, Sue, Nicola and Niki, and Nick for putting up with us. And everybody who was involved in Indiana Jones, my Magnum Opus.

Finally, thanks to Hazel, who has gotten me involved in so many Science Communication projects and helped me to find my future career.